Plusieurs approches en ondelettes pour la séparation et déconvolection de composantes. Application à des données astrophysiques.

Sandrine Anthoine

THÈSE

présentée pour obtenir le titre de

Docteur de l'École Polytechnique
Spécialité: Mathématiques Appliquées

par

Sandrine Anthoine

Approches en ondelettes pour la séparation et la déconvolution simultanées. Application à des données astrophysiques.

Soutenue le 5 août 2005, à Princeton University, jury composé de:

M.	Vincent Poor	Président
Mme.	Ingrid Daubechies	Directeur de thèse
M.	Stéphane Mallat	Co-directeur de thèse
Mme.	Elena Pierpaoli	Examinateur
M.	Ivan Selesnick	Rapporteur
M.	Peter Ramadge	Rapporteur

Abstract

This thesis addresses the problem of separating image components that have different structure, when different observations of blurred mixtures of these components are available. When only a single component is present and has to be extracted from a single observation, this reduces to the deblurring and denoising of one image, a problem well described in the image processing literature. On the other hand, the separation problem has been mainly studied in the simple case of linear mixtures (i.e. without blurring). In this thesis, the full problem is addressed globally, the separation being done simultaneously with the denoising and deblurring of the data at hand.

One natural way to tackle the multi-components/multi-observations problem in the blurred context is to generalize methods that exist for the enhancement of a single image. The first result presented in this thesis is a mathematical analysis of a heuristic iterative algorithm for the enhancement of a single image. This algorithm is proved to be convergent but not regularizing; a modification is introduced that restores this property. The main object of this thesis is to develop and compare two methods for the multi-components/multi-observations problem: the first method uses functional spaces to describe the signals; the second method models the local statistical properties of the signals. Both methods use wavelet frames to simplify the description of the data. In addition, the functional method uses different frames to characterize different components.

The performances of both algorithms are evaluated with regards to a particular astrophysical problem: the reconstruction of clusters of galaxies by the extraction of their Sunyaev-Zel'dovich effect in multifrequency measurements of the Cosmic Microwave Background anisotropies. Realistic simulations are studied, that correspond to different experiments, future or underway. It is shown that both methods yield clusters maps of sufficient quality for subsequent cosmological studies when the resolution of the observations is high and the level of noise moderate, that the noise level is a limiting factor for observations at lower resolution, and that the statistical algorithm is robust to the presence of point sources at higher frequencies.

Résumé

Cette thèse est consacrée au problème de séparation de composantes lorsque celles-ci sont des images de structure différente et que l'on en observe un ou plusieurs mélange(s) flou(s) et bruité(s). Les problèmes de déconvolution et de séparation, traditionnellement étudiés séparément, sont ici traités simultanément.

Une façon naturelle d'aborder le problème multicomposants/multiobservations est de généraliser les techniques de déconvolution d'une image unique. Le premier résultat présenté est une étude mathématique d'un tel algorithme. Preuve est faite que celui-ci est convergent mais pas régularisant et une modification restaurant cette propriété est proposée. Le sujet principal est le développement et la comparaison de deux méthodes pour traiter la déconvolution et séparation simultanées de composantes. La première méthode est basée sur les propriétés statistiques locales des composantes tandis que dans la seconde, ces signaux sont décrits par des espaces fonctionnels. Les deux méthodes utilisent des transformées en ondelettes redondantes pour simplifier les données.

Les performances des deux algorithmes sont évaluées et comparées dans le cadre d'un problème astrophysique : celui de l'extraction des amas de galaxies par l'effet Sunyaev-Zel'dovich dans les images multispectrales des anisotropies du fond cosmique. Des simulations réalistes sont étudiées. On montre qu'à haute résolution et niveau de bruit modéré, les deux méthodes permettent d'extraire des cartes d'amas de galaxies de qualité suffisante pour des études cosmologiques. Le niveau de bruit est un facteur limitant à basse résolution et la méthode statistique est robuste à la présence de points sources.

Contents

Présentation générale

Cette thèse a été préparée en cotutelle entre les laboratoires du Program in Applied and Computational Mathematics (PACM) à Princeton University (USA) et du Centre de Mathématiques APpliquées (CMAP) à l'École Polytechnique. Ce travail a été dirigé par le professeur Ingrid Daubechies et co-dirigé par le professeur Stéphane Mallat. Dans le cadre de la cotutelle, un unique manuscrit a été rédigé en anglais et ce présent chapitre constitue un résumé étendu en langue française. Il est à noter que ce chapitre est repris largement dans le chapitre 1 en anglais et que le lecteur à l'aise avec la langue anglaise peut donc commencer sa lecture au dit chapitre.

Le traitement des images

Les progrès technologique en matière de technique d'acquisition d'images ainsi qu'en terme de capacité de stockage de l'information sont à l'origine du fait qu'une masse colossale de données de plus en plus précises sont acquises dans l'espoir d'observer et comprendre des phénomènes de plus en plus fins. Il va donc de soi que les techniques de traitement d'images, c'est-à-dire les techniques qui servent à améliorer et analyser les images acquises doivent progresser en conséquence.

Le travail présenté dans cette thèse s'inscrit dans une optique d'analyse, de développement et d'évaluation de techniques mathématiques pour le traitement des images. L'analyse de techniques existantes permet de comprendre leur avantages et défauts pour développer des méthodes plus efficaces. Les méthodes développées ici le sont dans un cadre général mais leur évaluation se fait dans le cadre particulier d'une application en astrophysique. En effet, il est peu probable qu'une technique particulière soit bien adaptée à tout type d'images, une évaluation générale donne donc une idée imparfaite de la qualité des résultats obtenus en terme de la question scientifique à laquelle on souhaite répondre après traitement des images acquises. Le but de notre évaluation est donc détablir les performances des méthodes développées pour une application particulière, qui s'inscrit dans le cadre d'une collaboration avec des astrophysiciens et est à l'origine du développement de ces méthodes.

Cadre mathématique des problèmes abordés

Dans cette thèse, nous nous intéressons à des problèmes de traitement des données qui peuvent être décrits dans le cadre mathématique suivant. Nous cherchons à estimer

un ou plusieurs objets, notés $f_1, \ldots, f_M$, à partir d'une ou plusieurs observations, notées $g_1, \ldots, g_L$. Nous supposons que les processus d'acquisition des observations sont connus et peuvent être décrits par des opérateurs linéaires, à un terme d'erreur près. En d'autres termes, nous supposons connus les opérateurs linéaires $T_{m,l}$ tels que les observations g_l vérifient :

$$\forall l \in [\![1, L]\!], \quad g_l = \sum_{m=1}^{M} T_{m,l}\ f_m + n_l \tag{1.1}$$

où chaque terme n_l est un terme de bruit.

Ce cadre général permet de décrire des problèmes variés en traitement d'images. Parmi eux, on trouve les problèmes relatifs à l'amélioration d'une image unique ($M = L = 1$) tels que le débruitage ($T_{1,1}$ est l'identité) ou la déconvolution d'une image ($T_{1,1}$ représente une convolution). On trouve aussi les problèmes dit de fusion de données ($M = 1,\ L > 1$), où le même phénomène physique f_1 est observé grâce à différentes techniques : par exemple, une IRM du cerveau est enregistrée simultanément avec une électro-encéphalographie (EEG) de ce même cerveau, on obtient deux images g_1 et g_2 du même phénomène f_1 acquises sous différentes modalités, et la fusion de ces données consiste à utiliser les informations contenues dans ces deux acquisitions simultanées pour estimer le phénomène f_1. Enfin, on trouve aussi des problèmes où plusieurs phénomènes physiques se superposent dans les observations, il s'agit alors de séparer ces composantes.

Contributions

Les contributions de cette thèse se situent à plusieurs niveaux dans le cadre de l'étude et des problèmes décrits par l'équation (1.1).

Un permier volet de cette thèse est l'analyse mathématique d'un algorithme heuristique proposé pour la déconvolution d'une image. Cette analyse montre la convergence et identifie les conditions sous lesquelles cet algorithme est régularisant. Elle met aussi en évidence une propriété non-désirable de cet algorithme : il perd irrémédiablement de l'information dans certains cas. Nous proposons une légère modification qui garde les avantages de l'agorithme heuristique initial et ne présente plus ce défaut.

Dans un second volet, cette thèse présente deux méthodes de résolution de l'équation (1.1) adaptées aux cas où l'on souhaite réellement estimer plusieurs objets à partir de plusieurs observations ($M > 1$ et $L > 1$). L'une des méthodes est basée sur une description statistique locale des composantes à estimer et est adaptée au cas particulier de la déconvolution de mélanges de composantes. La seconde méthode passe par la minimisation d'une fonctionnelle variationnelle et permet de résoudre l'équation (1.1) dans le cadre général.

Enfin, ces deux méthodes sont mise en oeuvre et leurs performances sont comparées dans le cadre d'un problème astrophysique particulier : l'extraction des amas de galaxies à partir des données multifréquences d'observations du fond diffus cosmique. Cette étude prend en compte le fait que les caractéristiques (par exemple la résolution, le niveau de bruit...) varient grandement selon la mission d'observation astrophysique et nous évaluons les performances des algorithmes proposés en terme

de leur fiabilité pour les études astrophysique qui s'en suivent.

Plan du manuscrit

Après le chapitre 1 introductif, cette thèse est constituée de quatre chapitres. Les trois premiers sont théoriques et exposent les méthodes développées ainsi que l'étude d'un algorithme heuristique. Le dernier chapitre est dédiée à l'application astrophysique.

Plus précisément, le chapitre 2 est consacré à des méthodes de traitement de l'équation (1.1) par minimisation d'une fonctionnelle variationnelle. Le formalisme sur lequel nous nous basons est rappelé à la section 2.2. S'en suivent deux parties. La première consacrée à l'étude mathématique de l'algorithme heuristique de J-L Starck et de la modification proposée et fait l'objet des sections 2.3 et 2.4. La seconde partie décrit l'adaptation de la méthode variationnelle aux cas multi-objets/multi-observations et en particulier pour le problème de l'extraction des amas de galaxies et fait l'objet de la section 2.5.

Le chapitre 3 est également un chapitre théorique. Il décrit une méthode statistique pour traiter le problème posé par l'équation (1.1) dans le cas particulier de mélanges floues de composantes. Le modèle choisi pour décrire les composantes est expliqué dans la première section, la dérivation de l'estimateur dans la seconde section et les choix des différents paramètres dans la troisième section. La dernière section de ce chapitre explicite ce modèle dans le cadre de l'application à l'extraction des amas de galaxies.

Avant de passer à l'application astrophysique elle-même, nous rappelons au chapitre 4 les propriétés des systèmes d'ondelettes utilisés.

Le chapitre 5 détaille l'application des méthodes proposées à l'extraction des amas de galaxies à partir des données multifréquences d'observations du fond diffus cosmique. Les phénomèmes astrophysiques sont décrits dans la première section. Les méthodes d'estimation de la qualité des reconstructions font l'objet de la seconde section. Enfin les performances des algorithmes proposés sont comparées qualitativement et quantitativement dans le cadre de trois expériences aux spécifications différentes et les conclusions sont tirées dans la section finale.

Chapitre 1

Introduction

Imaging refers to the science of obtaining pictures or more complicated spatial representations, such as animations or 3-D computer graphics models, from physical objects. In a scientific context, the acquired images reflect measurements of physical quantities that are analyzed to understand the spatial properties of the observed phenomena. Imaging techniques have been developed to measure different quantities, with different resolution and reliability. These techniques keep improving, allowing us to collect and store more data, with greater precision, which in turns makes it possible to seek to understand finer scale phenomena. However, the quality of an image is naturally limited by the physical characteristics of the instrument used to collect the data, such as the size of the optical system and its maximum sampling rate, and by the physical limits linked to the phenomenon itself. E.g. the amplitude of the signal of interest may be very low compared to the amplitude of other signals that are necessarily imaged at the same time. Therefore image processing tools have to be developed simultaneously to imaging techniques, so that the improvements in image acquisition can be exploited optimally.

The contributions of this thesis are the analysis of existing methods and the development of new methods for the processing of images under the following assumptions : one seeks to recover the set of image components, $f_1, \ldots, f_M$, with $M \geq 1$, given a set of L observed images $g_1, \ldots, g_L$, with $L \geq 1$, knowing the linear operators $T_{m,l}$, $m \in [\![1, M]\!]$, $l \in [\![1, L]\!]$ such that the observed images g_l can be modeled by

$$\forall l \in [\![1, L]\!], \quad g_l = \sum_{m=1}^{M} T_{m,l} \; f_m + n_l \tag{1.1}$$

where each n_l denotes a noise term and $[\![k_1, k_2]\!]$ denotes the set : $\{k \in \mathbb{Z} : k_1 \leq k \leq k_2\}$. In this framework, the components $f_1, \ldots, f_M$ reflect measurements related to different phenomena. One may be interested in all, some, or even only one of them. A large set of image processing problems can be described by equation (1.1) : the denoising of one image ($M = L = 1$ and $T_{1,1}$ is the identity) ; the deblurring of one image ($M = L = 1$ and $T_{1,1}$ is a convolution) ; the fusion of images of the same phenomenon acquired by different modalities, ($M = 1$, $L > 1$) if the process of acquisition for each modality can be considered linear ; the extraction of components from several observations of linear mixtures of these ($M > 1$, $L > 1$, $T_{m,l}$ are scalars)...

There are many different ways to develop image processing algorithms. At one end of the spectra are algorithms giving the analytic solution to a mathematical problem where each unknown has been modeled precisely enough so that the solution is defined without ambiguity and can be computed. For example, if the image f and the noise n are independent Gaussian processes, then the conditional expectation of the random variable f given the random variable $g = f + n$, noted $E\{f|g\}$, is the best least-square estimate of f in the set g-measurable and square integrable random variables. That is, $E\{f|g\}$ the random variable $k(g)$ that minimizes the quantity $E\{|f - k(g)|^2\}$, with k measurable and $k(g)$ square integrable. If the covariance matrices $\mathbf{C_f}$ and $\mathbf{C_n}$ of f and n are known then $E\{f|g\}$ can be computed by the Wiener filter $E\{f|g\} = \mathbf{C_f}(\mathbf{C_f} + \mathbf{C_n})^{-1}g$. At the other end of the spectra are heuristic algorithms. These may give approximate solution to a well-defined mathematical problem that can not be solved analytically. More generally, heuristic algorithms are procedures designed to take advantage of some known properties of the signals, or to combine several aproaches, even when these are difficult to express mathematically. Unless an algorithm computes the analytic solution to a mathematical problem, its properties can only be studied experimentally.

The first contribution of this thesis is to provide a mathematical study of an adaptive iterative algorithm proposed by J-L. Starck in [58] to deconvolve one image. The algorithm proposed combines a known deblurring iterative scheme, with an adaptive projection on selected wavelet coefficients. This procedure was successfully used on astrophysical images, however, no mathematical study of this algorithm was provided. We review the mathematical framework proposed by I. Daubechies, M. Defrise and C. De Mol in [16] to solve inverse problems by another iterative algorithm in section 2.2, and show in section 2.3 how to use it to study J-L Starck's algorithm. We prove mathematically and by example that the proposed algorithm may give undesired results, namely that in the limit where the noise vanishes, the original image may not be recovered. In other words, this algorithm is not consistent. We propose a modification and show in section 2.4 that it restores consistency.

The deconvolution problem has been largely addressed in the literature in the case of a single image, i.e. when the problem is to restore the image f, from a blurred and noisy observation $g = T\ f + n = b * f + n$ ($*$ denotes the convolution). The task is not easy because the convolution operator is ill-conditioned, making it difficult to control the size of the noise term after inversion. A number of different algorithms have proposed, from simple linear filtering [63], to iterative algorithms [37, 49, 42], using deterministic [26] or statistical description of the data [33], and various tools such as PDE [51, 9] or multiscale decompositions [21, 30]... (see [34] for a more exhaustive list and description of deconvolution methods.) It has been established that deconvolution methods yield best results when the conditioning of the deconvolution operator and the structural properties of the image f and the noise n are taken into account at the same time ([30, 39]). The separation of different components, i.e. the estimation of M images $f_1, \ldots, f_M$ from linear mixtures ($g_l = \sum_{m=1}^{M} t_{m,l} f_m + n_l$, where the $t_{m,l}$ are scalars) has also been extensively studied [6, 61, 7]. Whether the scalars $t_{m,l}$ are assumed to be known or not, separating techniques seldom take into account the spatial properties of the different signals $f_1, \ldots, f_M$, or at least not to the same extent

as one does when processing a single image. This is harder to do in this context because the different properties of each component have to be handled at the same time. Both problems, the deconvolution and the separation, are usually studied independently of each other, and ad-hoc combinations are carried out if needed.

In this thesis, two new algorithms are proposed that simultaneously (denoise,) deblur and separate image components. More precisely, both algorithms compute estimates of the components $f_1, \ldots, f_M$ in Eq. (1.1) when each $T_{m,l}$ can be written $T_{m,l}(x) = a_{m,l}\ b_l * x$, where the $a_{m,l}$ are scalars and the $b_{m,l}$ are 2-dimensional point spread functions. The observations at hand are then modeled by :

$$\forall l \in [\![1, L]\!], \quad g_l = \sum_{m=1}^{M} a_{m,l}\ b_l * f_m + n_l. \tag{1.2}$$

Since the last equations can be rewritten : $\forall l \in [\![1, L]\!], \quad g_l = b_l * [\sum_{m=1}^{M} a_{m,l}\ f_m] + n_l$, the following two-steps algorithm seems like an appropriate solution : first deblur each g_l to obtain an estimate y_l of $\sum_{m=1}^{M} a_{m,l}\ f_m$ and secondly, separate the f_m from the y_l. However, in some cases it is desirable to avoid this intermediary step. This is the case for the extraction of clusters of galaxies from observations of Cosmic Microwave Background anisotropies, an application we study in detail in this thesis.

The observations g_l for this application are images of portions of the sky, obtained simultaneously at different light wavelengths (3 or 4 in the cases we considered). Each observed image is the convolution of the "true" image with a blurring beam function, which depends on the wavelength ; the observations are polluted by (Gaussian) noise that is independent from one image to another. The most intense components contained in the portion of sky observed besides the clusters of galaxies are the Cosmic Microwave Background (CMB) radiation, the Galaxy dust and infrared point sources. The contribution from each component to each observation depends on the wavelength. Hence the observations g_l can be modeled by equation (1.2) with $M = 4$ and $L = 3$ or 4. Our goal is to provide a "clean" image of the clusters of galaxies present in the observations, that will be usable by astrophysicists to derive properties of these clusters.

Clusters of galaxies are localized and compact objects sparsely distributed in the sky. The blurring by a beam function is especially badly conditioned at high frequencies, which correspond to small objects. Therefore, as mentioned earlier, the deconvolution of clusters of galaxies (supposing they were the only component present in the image), would be best when their localization is taken into account together with the properties of the convolution. Wavelet transforms are adapted to this situation because they are well localized both in frequency (and therefore constrain the conditioning of the convolution operator), and in space (so that clusters are well represented in wavelet space). However, in this case, the presence of other components complicates the task. The other components are much more intense than the clusters' signal, moreover they have very different spatial properties and the mixing scalar $a_{m,l}$ vary greatly with the frequencies of observation. Therefore the spatial properties of each intermediate deblurred image $y_l = \sum_{m=1}^{M} a_{m,l}\ f_m$ are different and do not reflect the properties of the clusters' signal. Since the latter is largely dominated in each y_l, it

would be very hard to recover a precise clusters' image using the two-steps technique proposed earlier. Instead, a method that solves the deconvolution and separation at the same time can exploit the fact that the same clusters' signal contributes to each observation and therefore should give better results.

We designed two different approaches to simultaneously deblur and separate image data. Both methods are flexible enough to take in account spatial properties that vary from one component to another. One method is based on a variational framework ; the other is more statistical in nature. The variational method uses a generalization of an algorithm proposed by I. Daubechies, M. Defrise and C. De Mol [16], that we explain and discuss in Chapter 2, sections 2.2 and 2.5. The method proposed is the minimization of the variational functional, by means of an iterative algorithm. In subsection 2.5.1, we describe how to this method solves the general problem posed by equation (1.1) (that is when the $T_{m,l}$ are general linear operators) and in the next subsection (2.5.2), we explain how to to use the method for our astrophysical application, deriving the parameters for separation of blurred mixtures and explaining how to model the properties of our astrophysical components. For the statistical approach, we were inspired by the work of J. Portilla, V. Strela, M. Wainwright and E. Simoncelli [48], which attacked the simultaneous denoising and deblurring of a single image. We explain in Chapter 3 how we extended this method to allow component separation (i.e. to solve Eq. (1.2)) and sketch the precise application to our astrophysical problem in Section 3.4.

As we noted earlier, the clusters' signal is well described in wavelet space. To avoid some drawbacks of the traditional decimated separable wavelet transform in two dimensions, we use different redundant wavelet transforms : the dual tree complex wavelet transforms for the variational approach [31, 32, 52, 53] and a steerable pyramid for the statistical approach (inspired by but not completely identical to the pyramid in [47]). The two transforms are described in Chapter 4, where we also discuss the algorithm we used to implement them.

Finally, in Chapter 5, we show and discuss the results of the two approaches on the astrophysical problem at hand, for several types of data sources. The resolution of data acquired previously is not sufficient to study the Sunyaev-Zel'dovich signature of clusters of galaxies, which is the particular effect we seek to estimate. However, several experiments are now being planned or underway, that will make it possible to do so. The different studies presented in Chapter 5 are made on realistic simulations of the data that will be acquired in the near future. (These simulations have been provided by astrophysicists.) This allows to assess the performances of both algorithms with respect to not only image processing standards but also with respect to the science that can be derived from these results. In particular, we asses the reliability in locating clusters of galaxies and the precision of the intensity estimated after extracting a cluster maps using both our algorithms. It turns out that each approach has strengths and weaknesses when compared to each other. A summary of these results is presented in Section 5.6.

Chapitre 2

Functional method

2.1 Framework

In this chapter, we consider the problem of deconvolution of mixtures of components as a variational problem, i.e. we wish to find estimates of the different components by minimizing a variational functional. We will consider functionals composed of a sum of discrepancy terms (one per observation) and regularization terms (one per component) :

$$J(f_1, f_2, \ldots, f_M) = \sum_{l=1}^{L} \rho_l \left\| (\sum_{m=1}^{M} T_{m,l} f_m - g_l) \right\|^2_{\mathcal{H}_l^o} + \sum_{m=1}^{M} \gamma_m |\!|\!| f_m |\!|\!|_{X_m} \ ; \tag{2.1}$$

here the $\mathcal{H}_l^o$ are Hilbert spaces, the γ_m and ρ_l are strictly positive scalars and the $|\!|\!|.|\!|\!|_{X_m}$ are norms. The observations at hand are the $\{g_l\}_{l\in[\![1,L]\!]}$. The $\{f_m\}_{m\in[\![1,M]\!]}$ are the components to be estimated. The mixing and blurring of component m at the frequency of observation number l is denoted by the linear operator $T_{m,l}$.

The minimizers of such a functional will strike a balance between the deviation of their image by the $T_{m,l}$ from the observed data on the one hand, and the $|\!|\!|.|\!|\!|_{X_m}$-norm on the other hand. This will give us a set of estimates $\widehat{f_1}, \widehat{f_2}, \ldots, \widehat{f_M}$ that have both properties of well approximating the observed data and having small $|\!|\!|.|\!|\!|_{X_m}$-norm. The $|\!|\!|.|\!|\!|_{X_m}$-norm here represent some "a priori knowledge" we have on the different components we are seeking : we expect the true component f_m to have a rather small $|\!|\!|.|\!|\!|_{X_m}$-norm. Note that the set of plausible images of one component, for example the set of CMB images, is not a vector space. So we do not try to design the vector space X_m so that each of its element corresponds to an image of component m. Rather, we design X_m so that the set of images of component m has a small $|\!|\!|.|\!|\!|_{X_m}$-norm. We hope that conversely, the estimate $\widehat{f_m}$ that we will obtain by minimizing (2.1) will be (close to) a plausible image of component m because it has a small $|\!|\!|.|\!|\!|_{X_m}$-norm. We shall use, for example, norms that penalize discontinuities or sharp transitions and norms that promote sparsity in a special representation like a wavelet representation. To do so, we embed the components f_m into Hilbert spaces $\mathcal{H}_m^i$ and consider $|\!|\!|.|\!|\!|_{X_m}$-norm

of the form :

$$|\!|\!| f |\!|\!|_{X_m} = \left[\sum_{\lambda \in \Lambda} w_\lambda^m |\, \langle f, \varphi_\lambda^m \rangle \,|^{p_m} \right]^{\frac{1}{p_m}} \tag{2.2}$$

where $\boldsymbol{\varphi}^m = \{\varphi_\lambda^m\}_{\lambda \in \Lambda}$ is a generating family of $\mathcal{H}_m^i$.

A general approach to solve problems of this nature can be found in [16, 14, 4]. The next section reviews the presentation in [16], which provides an iterative algorithm solving the problem when L=M=1. We then study two different generalizations. Section 2.3 and Section 2.4 are dedicated to the study of a slightly different problem where the discrepancy terms depend on the observation ; In Section 2.5, we generalize the iterative presented in [16] to solve the general case with M objects and L observations and describe its application to our astrophysical problem.

2.2 Iterative algorithm proposed by Daubechies, Defrise and De Mol

In this section, we summarize the findings presented in [16]. Daubechies, Defrise and De Mol present in this article an iterative algorithm to find a minimizer of Eq. (2.1) when $L = M = 1$. The goal is then to estimate a single object f_1 from a single observation g_1. To simplify the notations, we shall drop the indexes and denote $\mathcal{H}_1$ the Hilbert space of the object $\mathcal{H}_1^i$ and $\mathcal{H}_2$ the Hilbert space of the observation $\mathcal{H}_1^o$. The problem reduces to :

Problem 2.2.1. *Given $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$ an orthonormal basis of $\mathcal{H}_1$, a sequence of strictly positive weights $\boldsymbol{w}=\{w_\lambda\}_{\lambda\in\Lambda}$, a scalar $\gamma > 0$ and a scalar p with $1 \leq p \leq 2$, find :*

$$f^\star = \underset{f\in\mathcal{H}_1}{\operatorname{argmin}} \ \mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}(f) = \underset{f\in\mathcal{H}_1}{\operatorname{argmin}} \ \|Tf - g\|_{\mathcal{H}_2}^2 + \gamma |\!|\!| f |\!|\!|_{\mathbf{w},p}^p$$

where $|\!|\!| f |\!|\!|_{\mathbf{w},p} = \left[\sum_{\lambda\in\Lambda} w_\lambda |\, \langle f, \varphi_\lambda \rangle \,|^p\right]^{\frac{1}{p}} = \left[\sum_{\lambda\in\Lambda} w_\lambda |f_\lambda|^p\right]^{\frac{1}{p}}$.

Note that we used the notation $f_\lambda = \langle f, \varphi_\lambda \rangle$. We shall do so throughout this chapter unless specified otherwise.

The functional $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$ is convex, bounded below and verifies $\lim_{\|f\|\to\infty} \mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}(f) = +\infty$. Therefore it has a unique global minimum and has at least one minimizer. One can seek such a minimizer by canceling its partial derivative in f_λ :

$$\frac{\partial \mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}}{\partial f_\lambda}(f) = \ 2(T^*Tf)_\lambda - 2(T^*g)_\lambda + \gamma w_\lambda \ \operatorname{sign}(f_\lambda) \ |f_\lambda|^{p-1}.$$

If the operator T is the identity operator, then the equations decouple and the solution is given by solving $f_\lambda^\star = g_\lambda - \frac{\gamma w_\lambda}{2} \ \operatorname{sign}(f_\lambda) \ |f_\lambda|^{p-1}$. If $p = 1$, this reduces the the soft-thresholding operator (see [8]). However, when T is not the identity, these equations do not decouple which makes the problem harder to solve. Using surrogate functionals, one can define a sequence of similar problems that are easy to solve, and for which the sequence of minimizers obtained is strongly convergent in $\mathcal{H}_1$ to a solution of Problem 2.2.1. Moreover this scheme is regularizing. We explain it in detail below.

2.2.1 Surrogate functionals

Let us consider surrogate functionals $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}$ where a is an element of $\mathcal{H}_1$. The $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}$ are similar to $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$ but are slightly modified so that :

- For any a, $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}$ is strictly convex. Hence there exists a unique minimizer of $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}$, that we denote $f^{\star a}_{\gamma,\mathbf{w},p}$.
- The partial derivatives $\frac{\partial \mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}}{\partial f_\lambda}$ decouple. Therefore, one can find each coordinate $\{f^{\star a}_{\gamma,\mathbf{w},p}\}_\lambda$ independently by solving $\frac{\partial \mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}}{\partial f_\lambda} = 0$ for each λ.

Definition 2.2.2. *Given $a \in \mathcal{H}_1$ and C so that $\|\!|T^*T|\!\| < C$, the surrogate functional $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}} : \mathcal{H}_1 \to \mathbb{R}^+$ is defined by :*

$$\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}(f) = \|Tf - g\|^2_{\mathcal{H}_2} - \|Tf - Ta\|^2_{\mathcal{H}_2} + C\|f - a\|^2_{\mathcal{H}_1} + \gamma \|\!| f |\!\|^p_{\mathbf{w},p}$$

One verifies that the surrogate functional takes nonnegative values by noting that $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}(f) = \mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}(f) + C\|f - a\|^2_{\mathcal{H}_1} - \|Tf - Ta\|^2_{\mathcal{H}_2}$ with

$$\begin{aligned} C\|f - a\|^2_{\mathcal{H}_1} - \|Tf - Ta\|^2_{\mathcal{H}_2} &= C\|f - a\|^2_{\mathcal{H}_1} - \big\langle T(f-a),\ T(f-a)\big\rangle_{\mathcal{H}_2} \\ &= C\|f - a\|^2_{\mathcal{H}_1} - \big\langle f-a,\ T^*T(f-a)\big\rangle_{\mathcal{H}_1} \\ &\geq C\|f - a\|^2_{\mathcal{H}_1} - \|\!|T^*T|\!\| \|f-a\|^2_{\mathcal{H}_1} \\ &\geq (C - \|\!|T^*T|\!\|)\|f-a\|^2_{\mathcal{H}_1} \\ &\geq 0 \end{aligned}$$

Since $\|\!|T^*T|\!\| < C$, the term above is zero if and only if $f = a$, which ensures the strict convexity of the surrogate functional $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}$. Its partial derivatives in f_λ decouple :

$$\frac{\partial \mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}}{\partial f_\lambda}(f) = \ 2C\ f_\lambda - 2(C\ a + T^*g - T^*Ta)_\lambda + \gamma w_\lambda\ \mathrm{sign}(f_\lambda)\ |f_\lambda|^{p-1}.$$

and the minimizer of the surrogate functional $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}$ is :

$$\begin{aligned} f^{\star a}_{\gamma,\mathbf{w},p} &= \tfrac{1}{C}\,\mathbf{S}_{\gamma\mathbf{w},\mathrm{p}}\Big(C\ a + T^*g - T^*Ta\Big) \\ &= \tfrac{1}{C}\sum_\lambda S_{\gamma w_\lambda,p}\Big(\ \{\ C\ a + T^*g - T^*Ta\ \}_\lambda\ \Big)\ \varphi_\lambda \end{aligned} \tag{2.3}$$

Here,

$$S_{w,p}(x) \stackrel{def}{=} \Big(x + \frac{wp}{2}\ \mathrm{sign}(x)\ |x|^{p-1}\Big)^{-1}\ ,\ \text{for } 1 \leq p \leq 2, \tag{2.4}$$

where $(.)^{-1}$ denotes the inverse so that $S_{w,p}(x + \frac{wp}{2}\ \mathrm{sign}(x)\ |x|^{p-1}) = x$.

In particular, for $p = 1$, $S_{w,1}$ is the soft-thresholding operator :

$$S_{w,1}(x) = \begin{cases} x - w/2 & \text{if} \quad x \geq w/2 \\ 0 & \text{if} \quad |x| < w/2 \\ x + w/2 & \text{if} \quad x \leq -w/2 \end{cases} \tag{2.5}$$

Whereas for $p = 2$, one simply gets :

$$S_{w,2}(x) = \frac{x}{1+w} \tag{2.6}$$

The following proposition summarizes the properties of the surrogate functionals :

Proposition 2.2.3. *Suppose the operator T maps a Hilbert space $\mathcal{H}_1$ to another Hilbert space $\mathcal{H}_2$, with $|||T^*T||| < C$, and suppose g is an element of $\mathcal{H}_2$. Let $\{\varphi_\lambda\}_{\lambda\in\Lambda}$ be an orthonormal basis for $\mathcal{H}_1$, and let $\boldsymbol{w}=\{w_\lambda\}_{\lambda\in\Lambda}$ be a sequence of strictly positive numbers. Pick arbitrary $\gamma > 0$, $p \geq 1$ and $a \in \mathcal{H}_1$. Define the functional $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}(f)$ on $\mathcal{H}_1$ by*

$$\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}(f) = \|Tf - g\|^2_{\mathcal{H}_2} + \gamma \sum_{\lambda\in\Lambda} w_\lambda |f_\lambda|^p + C\|f - a\|^2_{\mathcal{H}_1} - \|T(f-a)\|^2_{\mathcal{H}_2} \ .$$

Then $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}(f)$ has a unique minimizer in $\mathcal{H}_1$.
This minimizer is given by $f = \frac{1}{C}\,\mathbf{S}_{\gamma\mathbf{w},\mathrm{p}}\,(Ca + T^*g - T^*Ta)$, where the operators $\mathbf{S}_{\mathbf{w},\mathrm{p}}$ are defined by

$$\mathbf{S}_{\mathbf{w},\mathrm{p}}(h) = \sum_\lambda S_{w_\lambda,p}(h_\lambda)\varphi_\lambda \quad , \tag{2.7}$$

with the functions $S_{w,p}$ from $\mathbb{R}$ to itself given by (2.4), (2.5) and (2.6).

Note that one can always assume that $C = 1$ since minimizing the surrogate functional $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}$ with the operator T and the observation g is the same problem as minimizing $\mathbf{J}^a_{\gamma,\mathbf{w}/C,\ p}$ with the operator $T' = \frac{1}{\sqrt{C}}T$, the observation $g' = \frac{1}{\sqrt{C}}g$ and the weights $\frac{\mathbf{w}}{C}$. This is also true for the initial functional $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$. Therefore, in the rest of this chapter, we will assume that $|||T^*T||| < 1$.

Next, we use a sequence of surrogate functionals and their minimizers to construct a solution of the original problem.

2.2.2 Iterative algorithm : convergence and stability

The iterative algorithm consists in minimizing a sequence of surrogate functionals $\mathbf{J}^{a^{\mathrm{n}}}_{\gamma,\mathbf{w},\mathrm{p}}(f)$, choosing a^n to be the minimizer obtained at the previous step :

Algorithm 2.2.4. *The iterative algorithm that solves Problem 2.2.1 proceeds as follows :*

$$\begin{cases} f^0 & \text{arbitrary} \\ f^n & = \underset{f\in\mathcal{H}_1}{\operatorname{argmin}} \left(\mathbf{J}^{f^{\mathrm{n-1}}}_{\gamma,\mathbf{w},\mathrm{p}}(f)\right) = \mathbf{S}_{\gamma\mathbf{w},\mathrm{p}}\,(f^{n-1} + T^*g - T^*Tf^{n-1}), \quad n \geq 1 \end{cases}$$

The two following theorems summarize the findings presented in [16]. The first theorem states that the iterative algorithm 2.2.4 converges strongly in the norm associated in the Hilbert space $\mathcal{H}_1$ for any initial guess f^0. The second theorem is concerned with the stability of the method. It gives sufficient conditions to ensure that the estimate recovered from a perturbed observation, $g = Tf_0 + e$, will approximate the object f_0 as the amplitude of the perturbation $\|e\|_{\mathcal{H}_2}$ goes to 0.

Theorem 2.2.5. *Let T be a bounded linear operator from $\mathcal{H}_1$ to $\mathcal{H}_2$, with norm strictly bounded by 1. Take $p \in [1,2]$, and let $\mathbf{S}_{\mathbf{w},\mathrm{p}}$ be the shrinkage operator defined by* (2.7)*, where the sequence $\boldsymbol{w}=\{w_\lambda\}_{\lambda\in\Lambda}$ is uniformly bounded below away from zero,*

i.e. there exists a constant $c > 0$ such that $\forall \lambda \in \Lambda : w_\lambda \geq c$. Then the sequence of iterates

$$f^n = \mathbf{S}_{\gamma \mathbf{w},\mathrm{p}} \left(f^{n-1} + T^* g - T^* T f^{n-1} \right) \ , \quad n = 1, 2, \dots \ ,$$

with f^0 arbitrarily chosen in $\mathcal{H}_1$, converges strongly to a minimizer of the functional

$$\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}(f) = \|Tf - g\|^2_{\mathcal{H}_2} + \gamma |\!|\!| f |\!|\!|^p_{\mathbf{w},p} \ ,$$

where $|\!|\!| f |\!|\!|_{\mathbf{w},p}$ denotes the norm $|\!|\!| f |\!|\!|_{\mathbf{w},p} = \left[\sum_{\lambda \in \Lambda} w_\lambda |\left\langle f, \varphi_\lambda \right\rangle|^p \right]^{1/p}$, $1 \leq p \leq 2$.

If the minimizer $f^\star$ of $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$ is unique, (which is guaranteed e.g. by $p > 1$ or $\ker(T) = \{0\}$), then every sequence of iterates f^n converges strongly to $f^\star$, regardless of the choice of f^0.

Theorem 2.2.6. *Assume that T is a bounded operator from $\mathcal{H}_1$ to $\mathcal{H}_2$ with $|\!|\!| T |\!|\!| < 1$, that $\gamma > 0$, $1 \leq p \leq 2$ and that the entries in the sequence $\boldsymbol{w} = \{w_\lambda\}_{\lambda \in \Lambda}$ are bounded below uniformly by a strictly positive number c. Assume that either $p > 1$ or $\ker(T) = \{0\}$. For any $g \in \mathcal{H}_2$ and any $\gamma > 0$, define $f^\star_{\gamma,\mathbf{w},p;g}$ to be the minimizer of $\mathbf{J}_{\gamma,\mathbf{w},p;g}(f)$. If $\gamma = \gamma(\epsilon)$ satisfies*

$$\lim_{\epsilon \to 0} \gamma(\epsilon) = 0 \qquad \text{and} \qquad \lim_{\epsilon \to 0} \frac{\epsilon^2}{\gamma(\epsilon)} = 0 \ , \tag{2.8}$$

then we have, for any $f_o \in \mathcal{H}_1$,

$$\lim_{\epsilon \to 0} \left[\sup_{\|g - T f_o\|_{\mathcal{H}_2} \leq \epsilon} \| f^\star_{\gamma(\epsilon),\mathbf{w},p;g} - f^\dagger \|_{\mathcal{H}_1} \right] = 0 \ ,$$

where $f^\dagger$ is the unique element of minimum $|\!|\!| \ |\!|\!|_{\mathbf{w},p}$–norm in the set $\mathcal{S}_{f_o} = \{f; Tf = Tf_o\}$.

2.2.3 Iterative algorithm with complex or redundant frames

The algorithms and theorems presented so far in this section apply only to the case where $\boldsymbol{\varphi} = \{\varphi_\lambda\}_{\lambda \in \Lambda}$ is an orthonormal basis of $\mathcal{H}_1$ and the scalar products $\langle ., \varphi_\lambda \rangle$ are real. It will be useful in our application to use redundant and/or complex families instead. To do that, one needs to make two changes, as was pointed out in [16].

Firstly, the definition of the operators $\mathbf{S}_{\mathbf{w},\mathrm{p}}$ has to be extended to complex numbers. This is done by applying $\mathbf{S}_{\mathbf{w},\mathrm{p}}$ only to the modulus of a complex number, keeping the phase fixed :

$$\mathbf{S}_{\mathbf{w},\mathrm{p}}(r.e^{i\theta}) \stackrel{def}{=} \mathbf{S}_{\mathbf{w},\mathrm{p}}(r).e^{i\theta}, \ r \in \mathbb{R}, \ \theta \in [0, 2\pi]. \tag{2.9}$$

This change is sufficient to prove Proposition 2.2.3 and Theorems 2.2.5 and 2.2.6 with the same algorithm 2.2.4.

Secondly, a clarification is required if the family $\boldsymbol{\varphi} = \{\varphi_\lambda\}_{\lambda \in \Lambda}$ is redundant. In that case, the set of sequences of scalar products of elements of $\mathcal{H}_1$:

$$\mathcal{C} = \left\{ \{\langle f, \varphi_\lambda \rangle\}_{\lambda \in \Lambda}, \ f \in \mathcal{H}_1 \right\},$$

is a strict subset of the set of square summable sequences $l^2(\mathbb{R})$ (or $l^2(\mathbb{C})$). As a consequence $f^{\star a}_{\gamma,\mathbf{w},p}$ defined in Eq.(2.3) need not be the minimizer of the surrogate functional $\mathbf{J}^a_{\gamma,\mathbf{w},\mathrm{p}}$ because

$$f = \tfrac{1}{C} \sum_{\lambda} S_{\gamma w_\lambda,p}\Big(\{ C\ a + T^*g - T^*Ta \}_\lambda \Big)\ \varphi_\lambda \tag{2.10}$$

does not imply that :

$$\forall \lambda,\ \langle f, \varphi_\lambda \rangle = \tfrac{1}{C} S_{\gamma w_\lambda,p}\Big(\{ C\ a + T^*g - T^*Ta \}_\lambda \Big) \tag{2.11}$$

In the derivation of algorithm 2.2.4, we used the fact that Eq. (2.10) and Eq. (2.11) are equivalent when $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$ is an orthonormal basis. When $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$ is redundant, this problem is rectified by projecting the sequence of coefficients obtained at each step of the iteration algorithm onto the set of scalar products $\mathcal{C}$:

$$f^n = P_{\mathcal{C}}\, \mathbf{S}_{\gamma\mathbf{w},\mathrm{p}} \left(f^{n-1} + T^*g - T^*Tf^{n-1} \right), \quad n \geq 1 \tag{2.12}$$

where $P_{\mathcal{C}}$ is the projection onto the set $\mathcal{C}$. (This can done more generally for any closed convex set $\mathcal{C}$, see Subsection 2.2.4.)

To illustrate the difference between a basis and a redundant frame, let us examine the case where the operator T is diagonal with respect to the tight frame $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$. That is, there exist scalars $\{t_\lambda\}_{\lambda\in\Lambda}$ such that :

$$\forall f \in \mathcal{H}_1,\ T(f) = T\Big(\sum_{\lambda\in\Lambda} \langle f, \varphi_\lambda \rangle\, \varphi_\lambda \Big) = \sum_{\lambda\in\Lambda} t_\lambda\, \langle f, \varphi_\lambda \rangle\, \varphi_\lambda. \tag{2.13}$$

We suppose that the algorithm is stopped after N steps.

If $\boldsymbol{\varphi}= \{\varphi_\lambda\}_{\lambda\in\Lambda}$ is an orthonormal basis, the iterations can be done in $l^2(\mathbb{R})$ (or $l^2(\mathbb{C})$) :

Algorithm 2.2.7. *First N steps of the iterative algorithm when T is diagonal on the orthonormal basis $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$:*

- *Pick f^o in $\mathcal{H}_1$ arbitrarily.*
- *Compute : $c^o_\lambda = \langle f^o, \varphi_\lambda \rangle$, $\forall \lambda \in \Lambda$.*
- *For $n = 1, .., N$, compute for all λ : $c^n_\lambda = S_{w,p}\Big((1 - t^2_\lambda) c^{n-1}_\lambda + t_\lambda g_\lambda \Big)$*
- *Output : $f^N = \sum_{\lambda\in\Lambda} c^N_\lambda \varphi_\lambda$.*

The intermediate estimates $f^1, \ldots, f^{N-1}$ need not be synthesized, only their frame coefficients, the c^n_λ, are computed. (For each n, $\{c^n_\lambda\}_{\lambda\in\Lambda}$ is a series in l^2.) We have : $c^n_\lambda = \langle f^n, \varphi_\lambda \rangle$, $\forall n$, $\forall \lambda$. Therefore, if $\boldsymbol{\varphi}= \{\varphi_\lambda\}_{\lambda\in\Lambda}$ is an orthonormal basis, one only needs to synthesize the final estimate f^N in $\mathcal{H}_1$, whereas if $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$ is redundant, one has to synthesize f^n at each step :

Algorithm 2.2.8. *First N steps of the iterative algorithm when T is diagonal on a redundant tight frame $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$:*

- *Pick $f^o \in \mathcal{H}_1$ arbitrary.*

- *Compute :* $c_\lambda^o = \langle f^o, \varphi_\lambda \rangle\,, \quad \forall \lambda \in \Lambda.$
- *For* $n = 1, .., N$, *compute :*
  - *For all* λ *:* $d_\lambda^n = S_{w,p}\big((1 - t_\lambda^2)c_\lambda^{n-1} + t_\lambda g_\lambda\big)$
  - $f^n = \sum_{\lambda \in \Lambda} d_\lambda^n \varphi_\lambda$
  - *For all* λ *:* $c_\lambda^n = \langle f^n, \varphi_\lambda \rangle$
- *Output :* f^N.

Note that because $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$ is redundant, although $\sum_{\lambda\in\Lambda} d_\lambda^n\varphi_\lambda = \sum_{\lambda\in\Lambda} c_\lambda^n\varphi_\lambda$, we do not have $d_\lambda^n = c_\lambda^n$. Therefore, one needs to synthesize f^n at each step to find the c_λ^n (this corresponds to the projection $P_{\mathcal{C}}$).

In the redundant case, f^n is not the minimizer of the surrogate functional at each step. The iterative algorithm still converges strongly. However, one can prove that the limit is the minimizer of the initial functional only in some cases. Generally though, it has been observed that using algorithm 2.2.4 yields good results with frames.

2.2.4 Iterative algorithm restricted to a closed convex set

The solution of problem 2.2.1 achieved by the iterative algorithm we presented is the minimizer of the functional $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$ in the whole Hilbert space $\mathcal{H}_1$. As explained in [16], it is possible to restrict the problem to a closed subset $\mathcal{D}$ of $\mathcal{H}_1$, for example the set of positive functions. The procedure consists in projecting the solution obtained at each step of the iterative algorithm onto the set $\mathcal{D}$:

$$f^n = P_{\mathcal{D}}\, \mathbf{S}_{\gamma\mathbf{w},\mathrm{p}} \left(f^{n-1} + T^*g - T^*Tf^{n-1}\right), \quad n \geq 1 \tag{2.14}$$

where $P_{\mathcal{D}}$ is the projection on the convex set $\mathcal{D}$. Some astrophysical components in our problem are positive and we will use this procedure to handle them.

Note that this is the same procedure that was used in the previous subsection to take in account the redundancy of the frame since the set of scalar products $\mathcal{C}$ is a closed subset of the set of square summable sequences.

2.3 Adaptive projections

In this section, we shall consider a generalization of the setting of [16], in which weights are introduced in the discrepancy term as well as in the prior. These weights were suggested originally by Jean-Luc Starck, in several papers and slightly different versions (see e.g. [58, 57, 43]). One of the algorithms suggested was :

Algorithm 2.3.1.

$$\left\{ \begin{array}{lcl} f^0 & \text{arbitrary} \\ f^n & = & \underset{f\in\mathcal{H}_1}{\operatorname{argmin}}\ \mathbf{S}_{\gamma,1}\left(f^{n-1} + T^*Mg - T^*MTf^{n-1}\right), \quad n \geq 1 \end{array} \right.$$

with $Mh = \sum_{\lambda\in\Lambda} m_\lambda h_\lambda \varphi_\lambda$, *and* $m_\lambda = 0$ *or* 1 *is chosen in function of* g_λ.

At first, it seems that the algorithm was purely heuristic, and was only later connected to a variational principle [59]. The weights m_λ in Starck's algorithm depend on the observation itself, and will make the analysis trickier ; we handle them by introducing an "adaptive projection operator".

2.3.1 Definition and corresponding iterative algorithm

Definition 2.3.2. *Given an orthonormal basis $\{\beta_\lambda\}_{\lambda\in\Lambda}$ of $\mathcal{H}_2$, an element g in $\mathcal{H}_2$ and a sequence of nonnegative thresholds $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$, the adaptive projection $\mathrm{M}_{\mathrm{g},\tau}$ is the map from $\mathcal{H}_2$ into itself defined by :*

$$\forall h \in \mathcal{H}_2, \quad \mathrm{M}_{\mathrm{g},\tau}(h) = \sum_{\lambda \ s.t. \ |g_\lambda|>\tau_\lambda} h_\lambda \beta_\lambda$$

(where, as usual, f_λ denotes the scalar product $\langle f, \beta_\lambda \rangle$)

Note that $\mathrm{M}_{\mathrm{g},\tau}$ is an orthogonal projection for any g and τ. It is therefore a continuous linear operator of unit norm, unless for all λ, $|g_\lambda| \leq \tau_\lambda$, in which case $\mathrm{M}_{\mathrm{g},\tau} = 0$. One can use the adaptive projection $\mathrm{M}_{\mathrm{g},\tau}$ to modify the similarity measure (discrepancy term) so that it discards the coordinates of the observation g that are deemed not reliable. More precisely, we consider in the fit to data term only the coordinate of index λ for which $|g_\lambda|$ is greater than some predefined value τ_λ. Problem 2.2.1 is thus modified into :

Problem 2.3.3. *Given a sequence of strictly positive weights $\boldsymbol{w}=\{w_\lambda\}_{\lambda\in\Lambda}$, a sequence of nonnegative thresholds $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$ and scalars γ and p with $\gamma > 0$ and $1 \leq p \leq 2$, find :*

$$f^\star = \underset{f\in\mathcal{H}_1}{\operatorname{argmin}} \ \mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}(f) = \underset{f\in\mathcal{H}_1}{\operatorname{argmin}} \ \| \mathrm{M}_{\mathrm{g},\tau}(Tf - g)\|^2_{\mathcal{H}_2} + \gamma |\!|\!| f |\!|\!|^p_{\mathbf{w},p}$$

where $|\!|\!| f |\!|\!|_{\mathbf{w},p}$ is defined in Problem 2.2.1 and $\mathrm{M}_{\mathrm{g},\tau}$ is defined above (2.3.2).

The value of the functional $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}(f)$ acting on operator T and observation g is exactly the value of the functional $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}(f)$ acting on operator $\mathrm{M}_{\mathrm{g},\tau}\, T$ and observation $\mathrm{M}_{\mathrm{g},\tau}\, g$. Hence once g and τ are fixed, Problem 2.3.3 is solved the same way as Problem 2.2.1 with the iterative algorithm modified accordingly :

Algorithm 2.3.4. *The iterative algorithm that solves Problem 2.3.3 proceeds as follows :*

$$\left\{ \begin{array}{ll} f^0 & \text{arbitrary} \\ f^n & = \mathbf{S}_{\gamma\mathbf{w},\mathrm{p}} \left(f^{n-1} + T^* \, \mathrm{M}_{\mathrm{g},\tau}\, g - T^* \, \mathrm{M}_{\mathrm{g},\tau}\, T f^{n-1} \right), \ n \geq 1 \end{array} \right.$$

Note that for $p = 1$, this is exactly the iterative algorithm 2.3.1 proposed by Jean-Luc Starck ! As is the case for Problem 2.2.1, the iterative algorithm 2.3.4 is strongly convergent in $\mathcal{H}_1$, regardless of the choice of f^0 and the limit is always a solution of Problem 2.3.3 :

Theorem 2.3.5. *Let T be a bounded linear operator from $\mathcal{H}_1$ to $\mathcal{H}_2$, with norm strictly bounded by 1. Take $p \in [1,2]$, $\{\tau_\lambda\}_{\lambda\in\Lambda}$ a sequence of nonnegative numbers and let $\mathbf{S}_{\mathbf{w},\mathrm{p}}$ be the shrinkage operator defined by (2.7), where the sequence $\{w_\lambda\}_{\lambda\in\Lambda}$ is uniformly bounded below away from zero, i.e. there $\exists c > 0$ s.t. $\forall\lambda \in \Lambda : \ w_\lambda \geq c$. Then the sequence of iterates*

$$f^n = \mathbf{S}_{\gamma\mathbf{w},\mathrm{p}} \left(f^{n-1} + T^* \, \mathrm{M}_{\mathrm{g},\tau} \, g - T^* \, \mathrm{M}_{\mathrm{g},\tau} \, T f^{n-1} \right) \ , \quad n = 1, 2, \dots \ ,$$

with f^0 arbitrarily chosen in $\mathcal{H}_1$, converges strongly to a minimizer of the functional

$$\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}(f) = \| \, \mathrm{M}_{\mathrm{g},\tau}(Tf - g)\|^2_{\mathcal{H}_2} + \gamma |\!|\!| f |\!|\!|^p_{\mathbf{w},p} \ ,$$

where $|\!|\!| f |\!|\!|_{\mathbf{w},p}$ denotes the norm $|\!|\!| f |\!|\!|_{\mathbf{w},p} = \left[\sum_{\lambda\in\Lambda} w_\lambda |\, \langle f, \varphi_\lambda \rangle \,|^p \right]^{1/p}$, $1 \leq p \leq 2$ and $\mathrm{M}_{\mathrm{g},\tau}(h) = \sum_{\lambda \ s.t. \ |g_\lambda|>\tau_\lambda} h_\lambda \beta_\lambda$.

If the minimizer $f^\star$ of $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}$ is unique, (which is guaranteed e.g. by $p > 1$ or $\ker(\mathrm{M}_{\mathrm{g},\tau} \, T) = \{0\}$), then every sequence of iterates f^n converges strongly to $f^\star$, regardless of the choice of f^0.

Démonstration. As we noted before :

$$\begin{array}{rcl} \mathbf{J}_{\gamma,\mathbf{w},p,\tau;T,g}(f) & = & \| \, \mathrm{M}_{\mathrm{g},\tau}(Tf - g)\|^2_{\mathcal{H}_2} \ + \ \gamma |\!|\!| f |\!|\!|^p_{\mathbf{w},p} \\ & = & \|(\mathrm{M}_{\mathrm{g},\tau} \, T)f - (\mathrm{M}_{\mathrm{g},\tau} \, g)\|^2_{\mathcal{H}_2} \ + \ \gamma |\!|\!| f |\!|\!|^p_{\mathbf{w},p} \\ & = & \mathbf{J}_{\gamma,\mathbf{w},p,0; \ \ T', \ g'}(f) \quad \text{with} \quad T' = \mathrm{M}_{\mathrm{g},\tau} \, T, \ \ g' = \mathrm{M}_{\mathrm{g},\tau} \, g \end{array}$$

Noting that $\mathbf{J}_{\gamma,\mathbf{w},p,0; \ \ T', \ g'}(f)$ is exactly the functional defined in Problem 2.2.1, it is then sufficient to prove $|\!|\!| T' |\!|\!|$ is strictly smaller than 1 to prove the strong convergence of the iterative algorithm 2.3.4 via Theorem 2.2.5. But $|\!|\!| T' |\!|\!| = |\!|\!| \, \mathrm{M}_{\mathrm{g},\tau} \, T |\!|\!| \leq |\!|\!| \, \mathrm{M}_{\mathrm{g},\tau} \, |\!|\!| . |\!|\!| T |\!|\!|$. Since $\mathrm{M}_{\mathrm{g},\tau}$ is an orthogonal projection, $|\!|\!| \, \mathrm{M}_{\mathrm{g},\tau} \, T |\!|\!| = 1$ or 0, and therefore $|\!|\!| T' |\!|\!| \leq |\!|\!| T |\!|\!| < 1$. ∎

2.3.2 Adaptive projections and diagonal operators

In this section, we illustrate the effects of the addition of the adaptive projection $\mathrm{M}_{\mathrm{g},\tau}$ in the iterative algorithm, by examining the simple case when T is a diagonal operator on the basis $\boldsymbol{\varphi} = \{\varphi_\lambda\}_{\lambda\in\Lambda}$: $Tf = \sum_{\lambda\in\Lambda} t_\lambda f_\lambda \varphi_\lambda$. In that case, the adaptive functional $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}$ reduces to :

$$\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}(f) = \sum_{\lambda\in\Lambda} \left(\delta_{\{|g_\lambda|>\tau_\lambda\}} (t_\lambda . f_\lambda - g_\lambda)^2 + \gamma \, w_\lambda \, |f_\lambda|^p \right) \tag{2.15}$$

Hence, the solution $f^\star$ is found by solving, independently for each λ :

$$f^\star_\lambda = \underset{x\in\mathbb{R}}{\mathrm{argmin}} \left(\delta_{\{|g_\lambda|>\tau_\lambda\}} (t_\lambda . x - g_\lambda)^2 + \gamma \, w_\lambda \, |x|^p \right) \tag{2.16}$$

If $|g_\lambda| \leq \tau_\lambda$ (or $t_\lambda = 0$), then $f^\star_\lambda = 0$, otherwise $f^\star_\lambda = S_{\gamma w_\lambda, p}(\, \overline{t_\lambda} . g_\lambda \,)$. Let us define the adaptive thresholding operator that maps $\mathbb{R}$ to itself by :

$$A_{\tau,\gamma,p}(x) = \begin{cases} S_{\gamma,p}(x) & \text{if } |x| > \tau \\ 0 & \text{otherwise} \end{cases} \tag{2.17}$$

Then, the solution of Eq. (2.15) is

$$f^\star = \sum_{\lambda \text{ s.t. } t_\lambda \neq 0} A_{\overline{t}\,\tau,\gamma,p}(\,\overline{t_\lambda}.g_\lambda\,)\;\varphi_\lambda. \tag{2.18}$$

This means that the introduction of the adaptive projection $\mathrm{M}_{g,\tau}$ results in combining a hard thresholding with parameter τ to the operator $S_{\gamma w_\lambda,p}$ when T is diagonal. The hard thresholding operator, or dead-zone function, maps $\mathbb{R}$ to itself and is defined by :

$$H_\tau(x) = \begin{cases} x & \text{if } |x| > \tau \\ 0 & \text{otherwise} \end{cases} \tag{2.19}$$

Suppose that T is the identity operator, that the weights $\{w_\lambda\}_{\lambda\in\Lambda}$ are identically 1 and that $p = 1$. If $\tau > \gamma$, the adaptive thresholding operator $A_{\tau,\gamma,1}$ (Fig.2.1, middle) is a compromise between the hard thresholding operator H_τ (Fig.2.1, left) and the soft-thresholding operator $S_{\gamma,1}$ (Fig.2.1 right) that would be used to solve Problem 2.2.1. (Note that if $\tau \leq \gamma$, the adaptive thresholding $A_{\tau,\gamma,1}$ reduces to the soft-thresholding $S_{\gamma,1}$).

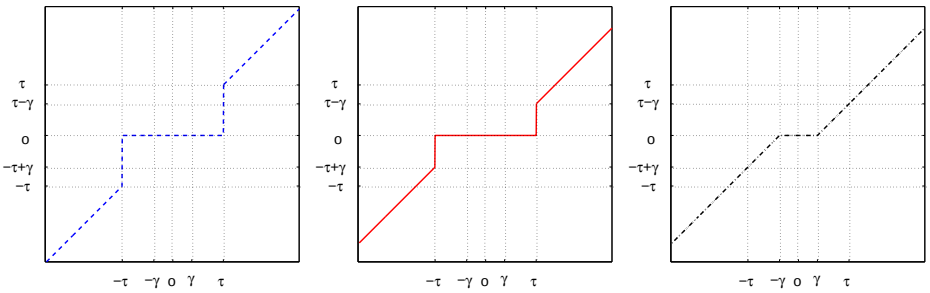

FIG. 2.1 – Left : hard thresholding operator H_τ ; middle : adaptive thresholding operator $A_{\tau,\gamma,1}$ right : soft-thresholding operator $S_{\gamma,1}$.

The hard thresholding operator H_τ can also be seen as an operator used for minimization :

$$H_\tau(g) = \underset{x\in\mathbb{R}}{\operatorname{argmin}}\ ((x-g).\delta_{\{|g|>\tau\}})^2 \tag{2.20}$$

Hence, H_τ corresponds to the limit of the adaptive thresholding operator $A_{\tau,\gamma,1}$ as γ goes to 0. On the other hand, the adaptive thresholding $A_{\tau,\gamma,1}$ is in fact the soft-thresholding $S_{\gamma,1}$ as soon as $\gamma > \tau$. It is therefore natural to examine the results of hard-thresholding, adaptive thresholding and soft-thresholding with a fixed value of τ so study the influence of γ. Fig. 2.2 displays such a study on a piecewise smooth signal. The top row of the figure shows the signal (left) and a noisy version of it (right) that is taken as the observation g. The signal is then reconstructed from g using adaptive, soft- or hard-thresholding with different values of the parameter γ for $\tau = 3$. The reconstructions obtained are displayed with γ increasing clockwise, i.e.

middle row left : $\gamma = 0$, middle row right : $\gamma = 1$, bottom row right : $\gamma = 2$, and bottom row left : $\gamma = 3$. The soft-thresholded reconstruction (bottom left) yields a smoother reconstruction than the hard-threshold (middle left) : the Gibbs effect is much weaker at the discontinuities of the signal for the soft-thresholding. But on the other hand it damps the signal, in particular the peaks. The adaptive thresholded reconstructions (middle right and bottom right) allow to find a different balance between the smoothness of the reconstruction and its precision for fast variations.

2.3.3 Stability

In this section, we investigate the regularization properties of the algorithm. Coarsely speaking, we would like the reconstructed components to very close to the true ones if the noise in the observation is negligible. More precisely, we will investigate whether $f_g^\star$ converges to f_o when $\|Tf_o - g\|_{\mathcal{H}_2}$ converges to zero. To do this, it will be convenient to first define some subsets of $\mathcal{H}_1$. The first subset, $\mathcal{M}_{f_o}$, is the set of elements of $\mathcal{H}_1$ that have the same image under T as f_o except maybe on the coordinates λ such that $(Tf_o)_\lambda = 0$:

Definition 2.3.6. *Given two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, an operator $T : \mathcal{H}_1 \to \mathcal{H}_2$, an orthonormal basis $\{\beta_\lambda\}_{\lambda\in\Lambda}$ of $\mathcal{H}_2$ and an element f_o of $\mathcal{H}_1$. The set $\mathcal{M}_{f_o}$ is the subset of elements of $\mathcal{H}_1$ that verify :*

$$f \in \mathcal{M}_{f_o} \iff M_{Tf_o,0}(Tf) = Tf_o \iff \Big[\{Tf_o\}_\lambda \neq 0 \Rightarrow \{Tf\}_\lambda = \{Tf_o\}_\lambda\Big]$$

For the coordinates λ such that $\{Tf_o\}_\lambda = 0$, one may have $\{Tf\}_\lambda \neq 0$ when f is in $\mathcal{M}_{f_o}$. If f_o is in $\ker(T)$ then $\mathcal{M}_{f_o} = \mathcal{H}_1$. On the contrary, if $\forall\lambda,\ \{Tf_o\}_\lambda \neq 0$, then $\mathcal{M}_{f_o}$ is exactly the subset of $\mathcal{H}_1$ having the same image as f_o under T. Note that $\mathcal{M}_{f_o}$ is closed and convex. We also define $\mathcal{H}_1^{T,\mathbf{w},p}$ as the set of elements f for which the corresponding set $\mathcal{M}_f$ has a unique minimizer for the $|\!|\!|.|\!|\!|_{\mathbf{w},p}$-norm.

Definition 2.3.7. *Given a Hilbert space $\mathcal{H}_1$, $\mathcal{H}_1^{T,\mathbf{w},p}$ is the subset of elements of $\mathcal{H}_1$ that verify : f_o is in $\mathcal{H}_1^{T,\mathbf{w},p}$ if and only if the set $\mathcal{M}_{f_o} = \{f : M_{Tf_o,0}Tf = Tf_o\}$ has a unique element of minimum $|\!|\!|.|\!|\!|_{\mathbf{w},p}$-norm.*

When $p > 1$, then $\mathcal{H}_1^{T,\mathbf{w},p} = \mathcal{H}_1$, regardless of T. This is not true if $p = 1$, even if $\ker T = \{0\}$. It turns out that algorithm 2.3.4 is regularizing for elements f in $\mathcal{H}_1^{T,\mathbf{w},p}$, and that the minimizer obtained in the limit $\|Tf_o - g\|_{\mathcal{H}_2}$ goes to zero is exactly the minimizer of the $|\!|\!|.|\!|\!|_{\mathbf{w},p}$-norm in $\mathcal{M}_{f_o}$. This is the object of the following theorem :

Theorem 2.3.8. *Assume that T is a bounded operator from $\mathcal{H}_1$ to $\mathcal{H}_2$ with $|\!|\!|T|\!|\!| < 1$, that $\gamma > 0$, $p \in [1, 2]$ and that the entries in the sequence $\boldsymbol{w}=\{w_\lambda\}_{\lambda\in\Lambda}$ are bounded below uniformly by a strictly positive number c.*

For any $g \in \mathcal{H}_2$ and any $\gamma > 0$ and any nonnegative sequence $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$, define $f^\star_{\gamma,\mathbf{w},p,\tau;g}$ to be a minimizer of $\mathbf{J}_{\gamma,\mathbf{w},p,\tau;g}(f)$. If $\gamma = \gamma(\epsilon)$ and $\tau = \tau(\epsilon)$ satisfy :

1. $\lim_{\epsilon\to 0} \gamma(\epsilon) = 0$

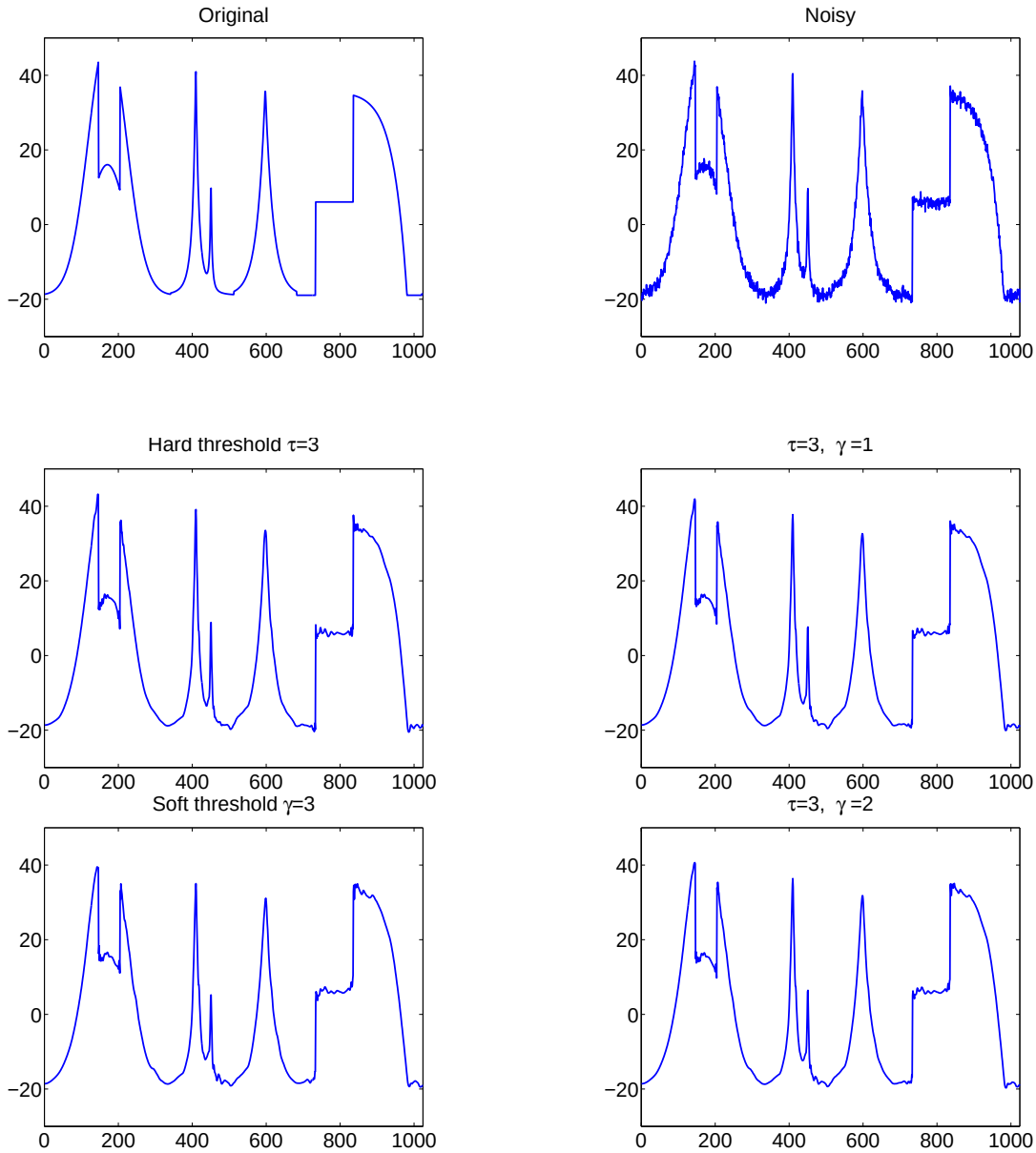

FIG. 2.2 – Top row, left : original signal ; right : noisy signal (white noise, $\sigma = 1$). Other rows : reconstructions with $\tau = 3$, increasing parameter γ clockwise ($\gamma = 0$ (hard-threshold), 1 , 2, 3 (soft-threshold)).

2. $\lim_{\epsilon\to 0} \dfrac{\epsilon^2}{\gamma(\epsilon)} = 0$

3. $\forall \lambda \in \Lambda,\ \lim_{\epsilon\to 0} \tau_\lambda(\epsilon) = 0$

4. $\exists\ \delta > 0, \quad s.t: \quad [\ \epsilon < \delta \Rightarrow \forall \lambda \in \Lambda,\ \tau_\lambda(\epsilon) > \epsilon\ ]$

then we have, for any $f_o \in \mathcal{H}_1^{T,\mathbf{w},p}$ *:*

$$\lim_{\epsilon\to 0}\left[\sup_{\|g-Tf_o\|_{\mathcal{H}_2}\leq\epsilon} \|f^\star_{\gamma(\epsilon),\mathbf{w},p,\tau(\epsilon);\ g} - f_o^\dagger\|_{\mathcal{H}_1}\right] = 0\ ,$$

where $f_o^\dagger$ *is the unique element of minimum* $\|\!|\ |\!\|_{\mathbf{w},p}$*–norm in the set* $\mathcal{M}_{f_o}$*.*

We will prove this stability theorem in a similar manner as Theorem 2.2.6 is proved in [16]. The proof proceeds as follows : first we prove that the norms $\|f^\star_{\gamma(\epsilon),\mathbf{w},p,\tau(\epsilon);\ g}\|_{\mathbf{w},p}$ are uniformly bounded. Secondly, we prove that when f_o is in $\mathcal{H}_1^{T,\mathbf{w},p}$, any sequence $\{f^\star_{\gamma(\epsilon_n),\mathbf{w},p,\tau(\epsilon_n);\ g_n}\}_n$ converges weakly to $f_o^\dagger$ when ϵ_n converges to 0. (Here g_n is any element in $\mathcal{H}_2$ verifying $\|g_n - Tf_o\|_{\mathcal{H}_2} \leq \epsilon_n$). Finally we prove strong convergence of the $\{f^\star_{\gamma(\epsilon_n),\mathbf{w},p,\tau(\epsilon_n);\ g_n}\}_n$ which proves Theorem 2.3.8.

Let us make some remarks before proving this theorem. One should point out the estimate $f_o^\dagger$ obtained through this algorithm is not necessarily what one expects. Indeed, even in the ideal case where T has a bounded linear inverse, we do not necessarily have $f_o^\dagger = f_o$. This can happen only when $\{Tf_o\}_\lambda = 0$ for some λ. If $\{Tf_o\}_\lambda \neq 0$ for all λ, then the projection $M_{Tf_o,0}$ is the identity and therefore $\mathcal{M}_{f_o} = \{f : M_{Tf_o,0}Tf = Tf_o\} = \{f : Tf = Tf_o\}$ and since T is one to one, this reduces to $\mathcal{M}_{f_o} = \{f_o\}$. This ensures that $f_o^\dagger = f_o$. However if $\{Tf_o\}_\lambda = 0$ for some λ, then $M_{Tf_o,0}$ is a projection with a non-trivial kernel : $\ker(M_{Tf_o,0}) = \mathrm{Span}\{\beta_\lambda,\ \lambda \text{ s.t. } \{Tf_o\}_\lambda = 0\}$. When the intersection : $\ker(M_{Tf_o,0}) \cap \mathrm{Im}(T)$ is not trivial i.e there exists some nonzero element h in $\mathcal{H}_1$ so that $Th_\lambda = Tf_{o\lambda}$ when $Tf_{o\lambda} \neq 0$, but Th_λ does not vanish for each λ where $Tf_{o\lambda} = 0$, then :

$$\{f_o\} \subsetneq \mathcal{M}_{f_o} = f_o + \ker(M_{Tf_o,0}) \cap \mathrm{Im}(T)$$

and therefore $f_o^\dagger$ need not be equal to f_o. This can happen even though T has a bounded linear inverse! Here is a simple example :

Example 1. *Consider* $T : \mathbb{R}^2 \to \mathbb{R}^2$*, the bounded and linear operator defined by :*

$$T : \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \ \mapsto\ \tfrac{1}{4}\begin{pmatrix} 2\ f_1 + f_2 \\ f_1 - f_2 \end{pmatrix} \quad \textit{and} \quad f_a = \begin{pmatrix} a \\ a \end{pmatrix} \textit{ for some } a \neq 0.$$

$\|\!|T|\!\| = \frac{1}{2} < 1$ *and although* T *has a bounded inverse :* $T^{-1} : \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \mapsto \frac{4}{3}\begin{pmatrix} f_1 + f_2 \\ f_1 - 2f_2 \end{pmatrix}$*, we have* $Tf_a = \begin{pmatrix} \frac{3a}{4} \\ 0 \end{pmatrix}$ *so that* $\mathcal{M}_{f_a} = \{f : (Tf)_1 = (Tf_a)_1\} = \{f : 2f_1 + f_2 = 3a\}$ *;*

The element in $\mathcal{M}_{f_a}$ *with minimal* l^1 *norm is :* $f_a^\dagger = \begin{pmatrix} \frac{3a}{2} \\ 0 \end{pmatrix}$*, and not* f_a *itself.*

Hence under the conditions of Theorem 2.3.8, solving Problem 2.3.3 will never enable us to recover f_a, even when we observe the unperturbed image Tf_a ! Indeed, in order to be stable, this algorithm has to discard the coordinates in $\mathcal{H}_2$ such that $Tf_{a\lambda} = 0$ even under an arbitrary small error of observation. The data-dependent truncation, introduced to find a more regular estimate when the noise is significant, looses the ability to recover f_a when its image is observed under ideal conditions.

We shall give more examples illustrating this peculiar behavior of the solutions to Problem 2.3.3 in the next subsection. But first, let us prove Theorem 2.3.8. To do so, we first examine the behavior of the projections $M_{g(\epsilon),\tau(\epsilon)}$ as ϵ goes to zero in the next two lemmas. The first lemma (Lemma 2.3.9) gives necessary and sufficient conditions on the sequence $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$ to that these projections converge in a weak sense as ϵ goes to zero. We will be interested in the case where the weak limit operator is $M_{Tf_o,0}$. The second lemma (Lemma 2.3.9) refines these conditions, so that in addition, the sequence $M_{g(\epsilon),\tau(\epsilon)}$ converges strongly to $M_{Tf_o,0}$ on the set : $T(\mathcal{M}_{f_0})$.

Lemma 2.3.9. *For $f \in \mathcal{H}_1$, let $\{g(\epsilon, f)\}_{\epsilon>0}$ be an arbitrary family of elements in $\mathcal{H}_2$ that satisfy $\|g(\epsilon, f) - Tf\|_{\mathcal{H}_2} < \epsilon$, $\forall\epsilon > 0$.*

1. *$\forall h \in \mathcal{H}_2$, $M_{g(\epsilon,f),\tau(\epsilon)}h$ converges weakly as ϵ goes to 0* ***if and only if*** *$\forall\lambda$: $\exists\ \delta(\lambda)$ such that either (a) or (b) holds, with*
   - *(a) $\forall\epsilon \in (0, \delta(\lambda))$, $\left|[g(\epsilon, f)]_\lambda\right| > \tau_\lambda$,*
   - *(b) $\forall\epsilon \in (0, \delta(\lambda))$, $\left|[g(\epsilon, f)]_\lambda\right| \leq \tau_\lambda$.*
2. *$M_{g(\epsilon,f),\tau(\epsilon)}$ converges weakly, independently of the choice of f and of the family $g(\epsilon, f)$, as ϵ goes to 0* ***if and only if*** *$\forall\lambda$: both (a) and (b) hold, with*
   - *(a) $\exists\ \delta(\lambda)$ such that $\forall\epsilon \in (0, \delta(\lambda))$, $\tau_\lambda(\epsilon) > \epsilon$*
   - *(b) $\lim_{\epsilon\to 0} \tau_\lambda(\epsilon) = 0$*

   In that case, the weak-limit operator is necessarily $\mathrm{M}_{Tf,0}$.
3. *When conditions 2.(a) and 2.(b) above hold, if $h(\epsilon)$ converges weakly to h, then $M_{g(\epsilon,f),\tau(\epsilon)}h(\epsilon)$ converges weakly to $\mathrm{M}_{Tf,0}\,h$ as ϵ goes to 0.*

Proof of Lemma 2.3.9. Let us examine the behavior of $M_{g(\epsilon,f),\tau(\epsilon)}$ coordinate by coordinate. Since $\left[M_{g(\epsilon,f),\tau(\epsilon)}h\right]_\lambda$ equals either h_λ or 0, depending on whether or not $\left|[g(\epsilon, f)]_\lambda\right| > \tau_\lambda(\epsilon)$, it follows that $M_{g(\epsilon,f),\tau(\epsilon)}(h)$ will converge weakly as ϵ goes to 0 if and only if for all coordinates λ, one of the following holds :

Either there exists some $\delta(\lambda) > 0$ such that $\left|[g(\epsilon, f)]_\lambda\right| > \tau_\lambda(\epsilon)$ for $\epsilon < \delta(\lambda)$. In this case, $\left[M_{g(\epsilon,f),\tau(\epsilon)}h\right]_\lambda = h_\lambda$ for $\epsilon < \delta(\lambda)$.

Or there exists some $\delta(\lambda) > 0$ such that $\left|[g(\epsilon, f)]_\lambda\right| \leq \tau_\lambda(\epsilon)$ for $\epsilon < \delta(\lambda)$. In this case, $\left[M_{g(\epsilon,f),\tau(\epsilon)}h\right]_\lambda = 0$ for $\epsilon < \delta(\lambda)$.

This proves the first assertion.

Let us now consider how uniform this behavior is in the choice of the family $g(\epsilon, f)$. Since $\left|[g(\epsilon, f) - Tf]_\lambda\right| \leq \|g(\epsilon, f) - Tf\|_{\mathcal{H}_2} \leq \epsilon$, the set of values that can be assumed by $|g(\epsilon, f)_\lambda|$ is exactly $\left[\ Tf - \epsilon, Tf + \epsilon\right]$ (take $g = Tf + r\beta_\lambda,\ r \in [-\epsilon, \epsilon]$ to reach all the values in this set). Therefore, for a fixed f, the weak convergence of the

operators $M_{g(\epsilon,f),\tau(\epsilon)}$, regardless of which sequence $g(\epsilon, f)$ is chosen, is equivalent to putting constraints on the sequence $\{\tau(\epsilon)_\lambda\}_{\lambda\in\Lambda}$ that depend of the coordinates $(Tf)_\lambda$. These constraints depends on whether $(Tf)_\lambda \neq 0$ or $(Tf)_\lambda = 0$:

- If $Tf_\lambda \neq 0$ then $\big\{|g(\epsilon, f)_\lambda|\big\} = \big[\ |Tf_\lambda| - \epsilon, |Tf_\lambda| + \epsilon\ \big]$. Therefore, one needs either : $\big[\epsilon < \delta(\lambda) \Rightarrow \tau_\lambda(\epsilon) > |Tf_\lambda| + \epsilon\big]$ or $\big[\epsilon < \delta(\lambda) \Rightarrow \tau_\lambda(\epsilon) \leq |Tf_\lambda| - \epsilon\big]$. In the first case, β_λ will always be in the kernel of $M_{g(\epsilon,f),\tau(\epsilon)}$ once $\epsilon < \delta(\lambda)$. In the second case β_λ will always in the range of $M_{g(\epsilon,f),\tau(\epsilon)}$ once $\epsilon < \delta(\lambda)$.
- If $Tf_\lambda = 0$ then $\{|g(\epsilon, f)_\lambda|\} = [0, \epsilon]$. Therefore one needs $[\epsilon < \delta(\lambda) \Rightarrow \tau_\lambda(\epsilon) > \epsilon]$. In this case, β_λ will always be in the kernel of $M_{g(\epsilon,f),\tau(\epsilon)}$ once $\epsilon < \delta(\lambda)$.

Note that we do not know beforehand the value of Tf. To be useful, we must derive requirements on the parameters $\tau_\lambda(\epsilon)$ that do not depend on f. The minimum requirements on $\tau(\epsilon)$ ensuring the operators $M_{g(\epsilon,f),\tau(\epsilon)}$ converge weakly as ϵ goes to 0 are :

- $\forall\lambda,\ \lim_{\epsilon\to 0} \tau_\lambda(\epsilon) = 0$: this ensures that if $Tf_\lambda \neq 0$, we will have $\tau_\lambda(\epsilon) < |Tf_\lambda| - \epsilon$ for sufficiently small ϵ.
- $\forall\lambda,\ \exists\delta(\lambda)$ such that $\epsilon < \delta(\lambda) \Rightarrow \tau_\lambda(\epsilon) < \epsilon$: this ensures that if $Tf_\lambda = 0$, we will have $\tau_\lambda(\epsilon) < |Tf_\lambda| + \epsilon = \epsilon$ for sufficiently small ϵ.

If these conditions are satisfied, the $M_{g(\epsilon,f),\tau(\epsilon)}$ converge weakly as ϵ goes to 0 and one can determine the weak limit :

- for λ s.t. $Tf_\lambda \neq 0$: $\lim_{\epsilon\to 0} \tau_\lambda(\epsilon) = 0$ hence there exists $\delta(\lambda, f)$ such that $\epsilon < \delta(\lambda, f)$ implies $\tau_\lambda(\epsilon) < |Tf_\lambda| - \epsilon$. It follows that : $|g(\epsilon, f)_\lambda| > \tau_\lambda(\epsilon)$ so that $M_{g(\epsilon,f),\tau(\epsilon)}(\beta_\lambda) = \beta_\lambda$ for any $g(\epsilon, f)$ and any $\epsilon < \delta(\lambda, f)$
- for λ s.t. $Tf_\lambda = 0$: $\epsilon < \delta(\lambda)$ implies $\tau_\lambda(\epsilon) > \epsilon$. It follows that if $\epsilon < \delta(\lambda)$, then $|g(\epsilon, f)_\lambda| > \tau_\lambda(\epsilon)$ so that $M_{g(\epsilon,f),\tau(\epsilon)}(\beta_\lambda) = 0$ for any $g(\epsilon, f)$ and any $\epsilon < \delta(\lambda)$.

This proves that the weak limit of $M_{g(\epsilon,f),\tau(\epsilon)}$ for any fixed f is $\mathrm{M}_{Tf,0}$ and finishes the proof of the second part of Lemma 2.3.9.

Finally, assuming $h(\epsilon)$ converges weakly to h, we have $\forall\lambda$:

$$\begin{aligned} & \Big|\big[M_{g(\epsilon,f),\tau(\epsilon)}h(\epsilon) - \mathrm{M}_{Tf,0}\,h\ \big]_\lambda\Big| && (2.21)\\ = & \Big|\big[M_{g(\epsilon,f),\tau(\epsilon)}(h(\epsilon) - h) + (M_{g(\epsilon,f),\tau(\epsilon)} - \mathrm{M}_{Tf,0})h\ \big]_\lambda\Big| && (2.22)\\ = & \Big|\big[M_{g(\epsilon,f),\tau(\epsilon)}(h(\epsilon) - h)\ \big]_\lambda\Big| + \Big|\big[M_{g(\epsilon,f),\tau(\epsilon)}h - \mathrm{M}_{Tf,0}\,h\ \big]_\lambda\Big| && (2.23)\end{aligned}$$

The second term vanishes as ϵ goes to 0 because $M_{g(\epsilon,f),\tau(\epsilon)}$ converges weakly to $\mathrm{M}_{Tf,0}$ when the conditions 2.(a) and 2.(b) hold. Moreover, we have seen in the proof of the second part of the lemma that for any λ :

- either there exists a $\delta(\lambda)$ such that $M_{g(\epsilon,f),\tau(\epsilon)}(\beta_\lambda) = 0$ for any $\epsilon < \delta(\lambda)$. In that case, $\Big|\big[M_{g(\epsilon,f),\tau(\epsilon)}(h(\epsilon) - h)\ \big]_\lambda\Big| = 0$, for $\epsilon < \delta(\lambda)$.
- or there exists a $\delta(\lambda)$ such that $M_{g(\epsilon,f),\tau(\epsilon)}(\beta_\lambda) = \beta_\lambda$ for any $\epsilon < \delta(\lambda)$. In that case, $\Big|\big[M_{g(\epsilon,f),\tau(\epsilon)}(h(\epsilon) - h)\ \big]_\lambda\Big| = \Big|\big[h(\epsilon) - h\ \big]_\lambda\Big|$, for $\epsilon < \delta(\lambda)$; and the weak convergence of $h(\epsilon)$ to h allows to conclude that $\Big|\big[M_{g(\epsilon,f),\tau(\epsilon)}(h(\epsilon) - h)\ \big]_\lambda\Big| \to 0$

This proves that $M_{g(\epsilon,f),\tau(\epsilon)}h(\epsilon)$ converges weakly to $\mathrm{M}_{Tf,0}\,h$ and finishes the proof of Lemma 2.3.9. ∎

We shall now see how to ensure strong convergence of the $M_{g(\epsilon,f),\tau(\epsilon)}(h)$ when h is in $\mathcal{M}_f$.

Lemma 2.3.10. *If there exists a value of δ independent of λ such that $\forall \epsilon < \delta$ and $\forall \lambda, \quad \tau_\lambda(\epsilon) > \epsilon$, then the two following properties hold :*

1. *For any choice of f and of the family $g(\epsilon, f)$:*

$$\forall \epsilon < \delta, \ M_{g(\epsilon,f),\tau(\epsilon)} = M_{Tf,0} M_{g(\epsilon,f),\tau(\epsilon)} = M_{g(\epsilon,f),\tau(\epsilon)} M_{Tf,0} = \sum_{\substack{\lambda \ s.t. \ Tf_\lambda \neq 0 \\ and \ |g_\lambda| \geq \tau_\lambda}} \langle \ ., \beta_\lambda \rangle \, \beta_\lambda.$$

2. *In particular, for any choice of $f \in \mathcal{H}_1^{T,\mathbf{w},p}$ and of the family $g(\epsilon, f)$, (i.e. whenever $\mathcal{M}_f$ has a unique minimizer $f^\dagger$ of the $|\!|\!|.|\!|\!|_{\mathbf{w},p}$-norm) :*

$$\forall \epsilon < \delta, \quad M_{g(\epsilon,f),\tau(\epsilon)}(Tf^\dagger) = M_{g(\epsilon,f),\tau(\epsilon)}(Tf).$$

Proof of Lemma 2.3.10 : The first part of Lemma 2.3.10 results from properties of orthogonal projections. If P_1 and P_2 are two orthogonal projections, then :

$$\begin{array}{rcl} P_1 \ P_2 & = & P_2 \ P_1 \\ \ker(P_2) \subset \ker(P_1) & \Leftrightarrow & P_1 P_2 = P_1. \end{array}$$

Hence, we already proved $M_{g(\epsilon,f),\tau(\epsilon)} \ M_{Tf,0} = M_{Tf,0} \ M_{g(\epsilon,f),\tau(\epsilon)}$ and

$$M_{g(\epsilon,f),\tau(\epsilon)} M_{Tf,0} = M_{g(\epsilon,f),\tau(\epsilon)} \Leftrightarrow \left[(Tf)_\lambda = 0 \Rightarrow |g_{(\epsilon,f)_\lambda}| \leq \tau_\lambda(\epsilon) \right].$$

When f and ϵ are fixed, the right hand side holds for any $g(\epsilon, f)$ if and only if $\left[(Tf)_\lambda = 0 \Rightarrow \epsilon < \tau_\lambda(\epsilon) \right]$ which proves the first part of Lemma 2.3.10.

For f in $\mathcal{H}_1^{T,\mathbf{w},p}$, $f^\dagger$ is well defined and verifies $M_{Tf,0} Tf^\dagger = Tf$. Applying $M_{g(\epsilon,\tau(\epsilon))}$ to this equality and using the previous result finishes the proof of Lemma 2.3.10. ∎

With the help of these two lemma, we can now proceed to the

Proof of Theorem 2.3.8 : Let us consider f_o in $\mathcal{H}_1^{T,\mathbf{w},p}$, i.e. f_o verifies that $\mathcal{M}_{f_o}$ has a unique minimizer $|\!|\!|\ |\!|\!|_{\mathbf{w},p}$–norm. We note this minimizer $f_o^\dagger$. We fix the following sequences : $\{\epsilon_n\}_n$ such that $\epsilon_n \underset{n\to\infty}{\longrightarrow} 0$, $\{g_n\}_n$ such that $\forall n, \ \|g_n - Tf_o\|_{\mathcal{H}_2} \leq \epsilon_n$, and $\{\gamma_n\}_n \overset{def}{=} \{\gamma(\epsilon_n)\}_n$ and $\{\tau_n\}_n \overset{def}{=} \{\tau(\epsilon_n)\}_n$ that verify conditions 1 to 4 in Theorem 2.3.8. For every n, we choose a minimizer $f_n^\star \overset{def}{=} f^\star_{\gamma_n,\mathbf{w},p,\tau_n;\, g_n}$ of the functional $J_n(f) \overset{def}{=} J_{\gamma_n,\mathbf{w},p,\tau_n;\, g_n}(f) = \| \mathrm{M}_{g_n,\tau_n}(Tf - g_n)\|^2_{\mathcal{H}_2} + \gamma_n |\!|\!| f |\!|\!|^p_{\mathbf{w},p}$.

We want to prove that for any such choice of the ϵ_n, g_n, γ_n, τ_n and $f_n^\star$, the sequence $f_n^\star$ converges strongly in $\mathcal{H}_1$ to $f_o^\dagger$, where $f_o^\dagger$ is the unique minimizer of the $|\!|\!|.|\!|\!|_{\mathbf{w},p}$-norm in the set $\mathcal{M}_{f_o} = \{f : (Tf)_\lambda = (Tf_o)_\lambda, \ \forall \lambda \ s.t. \ (Tf_o)_\lambda \neq 0\}$. We will also note $M_n \overset{def}{=} \mathrm{M}_{g_n,\tau_n}$.

The sequences $\{|||f_n^\star|||_{\mathbf{w},p}\}_n$ and $\{\|f_n^\star\|_{\mathcal{H}_1}\}_n$ are uniformly bounded :

By definition of J_n, $\forall n$:

$$\begin{aligned} |||f_n^\star|||_{\mathbf{w},p}^p &\leq \tfrac{1}{\gamma_n} J_n(f_n^\star) \\ \text{so that} \quad |||f_n^\star|||_{\mathbf{w},p}^p &\leq \tfrac{1}{\gamma_n} J_n(f_o^\dagger) \quad \text{since } f_n^\star \text{ minimizes } J_n. \end{aligned}$$

But :

$$\begin{aligned} J_n(f_o^\dagger) &= \|M_n(Tf_o^\dagger - g_n)\|_{\mathcal{H}_2}^2 + \gamma_n |||f_o^\dagger|||_{\mathbf{w},p}^p \\ &\leq \|M_n(Tf_o^\dagger - Tf_o)\|_{\mathcal{H}_2}^2 + \|M_n(Tf_o - g_n)\|_{\mathcal{H}_2}^2 + \gamma_n |||f_o^\dagger|||_{\mathbf{w},p}^p \\ &\leq \|M_n(Tf_o^\dagger - Tf_o)\|_{\mathcal{H}_2}^2 + |||M_n|||^2 . \|(Tf_o - g_n)\|_{\mathcal{H}_2}^2 + \gamma_n |||f_o^\dagger|||_{\mathbf{w},p}^p \\ &\leq \|M_n(Tf_o^\dagger - Tf_o)\|_{\mathcal{H}_2}^2 + \epsilon_n^2 + \gamma_n |||f_o^\dagger|||_{\mathbf{w},p}^p \end{aligned}$$

where we used $|||M_n|||^2 \leq 1$ and $\|Tf_o - g_n\| \leq \epsilon_n$ in the last equation. Hence

$$\forall n, \ |||f_n^\star|||_{\mathbf{w},p}^p \leq \frac{\|M_n(Tf_o^\dagger - Tf_o)\|_{\mathcal{H}_2}^2}{\gamma_n} + \frac{\epsilon_n^2}{\gamma_n} + |||f_o^\dagger|||_{\mathbf{w},p}^p. \tag{2.24}$$

Since condition 3 and 4 of Theorem 2.3.8 are satisfied, we can use Lemma 2.3.10.(2). It follows that if n is large enough, $M_n T f_o^\dagger = M_n T f_o$. Moreover, $\frac{\epsilon_n^2}{\gamma_n} \underset{n\to\infty}{\longrightarrow} 0$ by condition 2 of Theorem 2.3.8. This proves that $\{|||f_n^\star|||_{\mathbf{w},p}\}_n$ is uniformly bounded.

Since $\mathbf{w}$ is bounded below by $c > 0$ and $p \leq 2$, the $\|.\|_{\mathcal{H}_1}$-norm is bounded above by $c^{-\frac{1}{p}} |||.|||_{\mathbf{w},p}$:

$$|f_\lambda| = \big(|f_\lambda|^p\big)^{\frac{1}{p}} \leq \big(\tfrac{w_\lambda}{c}|f_\lambda|^p\big)^{\frac{1}{p}} \leq \big(\sum_{\lambda\in\Lambda} \tfrac{w_\lambda}{c}|f_\lambda|^p\big)^{\frac{1}{p}} = c^{-\frac{1}{p}} |||f|||_{\mathbf{w},p} \tag{2.25}$$

so that :

$$\|f\|_{\mathcal{H}_1}^2 = \sum_{\lambda\in\Lambda} |f_\lambda|^2 \leq \sum_{\lambda\in\Lambda} \tfrac{w_\lambda}{c}|f_\lambda|^p |f_\lambda|^{2-p} \leq \sum_{\lambda\in\Lambda} \tfrac{w_\lambda}{c}|f_\lambda|^p \big[c^{-\frac{1}{p}} |||f|||_{\mathbf{w},p}\big]^{2-p} \tag{2.26}$$

$$\|f\|_{\mathcal{H}_1}^2 \leq \tfrac{1}{c} |||f|||_{\mathbf{w},p}^p \big[c^{-\frac{1}{p}} |||f|||_{\mathbf{w},p}\big]^{2-p} = c^{-\frac{2}{p}} |||f|||_{\mathbf{w},p}^2 \tag{2.27}$$

Hence, the sequence $\{f_n^\star\}$ is also uniformly bounded in $\mathcal{H}_1$.

$f_o^\dagger$ is the unique accumulation point of the sequence $\{f_n^\star\}_n$:

Since it is uniformly bounded in $\mathcal{H}_1$, the sequence $\{f_n^\star\}_n$ has at least one weakly convergent subsequence $\{f_k^\star\}_k$. Let us denote its weak limit $\widetilde{f}$. We shall now prove that $\widetilde{f} = f_o^\dagger$.

Since $f_k^\star$ is a minimizer of J_k obtained through the iterative algorithm, 2.3.4, it verifies the fixed point equation : $f_k^\star = \mathbf{S}_{\gamma_k \mathbf{w},p}\,(f_k^\star + T^* M_k g_k - T^* M_k T f_k^\star)$. We note

$h_k = f_k^\star + T^* M_k g_k - T^* M_k T f_k^\star$, so that $f_k^\star = \mathbf{S}_{\gamma_k \mathbf{w},\mathrm{p}}(h_k)$. By definition of the weak limit, it follows that :

$$\begin{aligned}
\forall \lambda,\ \widetilde{f}_\lambda &= \lim_{k\to\infty} S_{\gamma_k w_\lambda}\big((h_k)_\lambda\big) \\
&= \lim_{k\to\infty} [(h_k)_\lambda] + \lim_{k\to\infty} \left[S_{\gamma_k w_\lambda}\left((h_k)_\lambda\right) - (h_k)_\lambda\right] && \text{but } \lim_{k\to\infty} \gamma_k w_\lambda = 0 \\
\text{So, } \forall \lambda,\ \widetilde{f}_\lambda &= \lim_{k\to\infty} [(h_k)_\lambda] && \text{since } \forall x,\ S_v(x) \xrightarrow[v\to 0]{} x \\
&= \lim_{k\to\infty} [(f_k^\star + T^* M_k g_k - T^* M_k T f_k^\star)_\lambda] \\
&= \widetilde{f}_\lambda + \lim_{k\to\infty} [(T^* M_k g_k - T^* M_k T f_k^\star)_\lambda] && \text{since } (f_k^\star)_\lambda \xrightarrow[k\to\infty]{} \widetilde{f}_\lambda.
\end{aligned}$$

As a result : $\forall \lambda,\ \lim_{k\to\infty} [(T^* M_k g_k - T^* M_k T f_k^\star)_\lambda] = 0$.

But since $\|g_k - Tf_o\|_{\mathcal{H}_2} \leq \epsilon_k$, then $\|T^* M_k (g_k - Tf_o)\|_{\mathcal{H}_1} \leq |\!|\!| T^* |\!|\!| |\!|\!| M_k |\!|\!| \epsilon_k < \epsilon_k$. This proves that for all λ :

$$\lim_{k\to\infty} [(T^* M_k T f_o - T^* M_k T f_k^\star)_\lambda] = 0. \tag{2.28}$$

Moreover, from Lemma 2.3.9.(2), we know that $\{M_k(Tf_o)\}_k$ converges weakly to $M_{Tf_o,0}(Tf_o) = Tf_o$. Together with the continuity of T^*, this leads to :

$$T^* M_k T f_k^\star \xrightarrow[k\to\infty]{w} T^* T f_o. \tag{2.29}$$

On the other hand, $f_k^\star$ converges weakly to $\widetilde{f}$. Using the continuity of T, we get $Tf_k^\star \xrightarrow[k\to\infty]{w} T\widetilde{f}$. From Lemma 2.3.9.(3), this also implies $\{M_k T f_k^\star\}_k \xrightarrow[k\to\infty]{w} \mathrm{M}_{Tf_o,0}\, T\widetilde{f}$. and it follows from the continuity of T^* that :

$$T^* M_k T f_k^\star \xrightarrow[k\to\infty]{w} T^* M_{Tf_o,0} T\widetilde{f}. \tag{2.30}$$

Plugging this last result in Eq. (2.29), we obtain the equality :

$$T^* M_{Tf_o,0} T\widetilde{f} = T^* T f_o \tag{2.31}$$

Since $M_{Tf_o,0}(Tf_o) = Tf_o$, the previous equality reduces to : $T^* M_{Tf_o,0} T(\widetilde{f} - f_o) = 0$. Taking the scalar product with $\widetilde{f} - f_o$, we obtain :

$$\begin{aligned}
\left\langle \widetilde{f} - f_o,\ T^* M_{Tf_o,0} T(\widetilde{f} - f_o) \right\rangle &= 0 \\
&\Leftrightarrow \left\langle M_{Tf_o,0} T(\widetilde{f} - f_o),\ M_{Tf_o,0} T(\widetilde{f} - f_o) \right\rangle = 0 \\
&\Leftrightarrow \|M_{Tf_o,0} T(\widetilde{f} - f_o)\|^2_{\mathcal{H}_2} = 0 \\
&\Leftrightarrow M_{Tf_o,0} T(\widetilde{f} - f_o) = 0 \\
&\Leftrightarrow M_{Tf_o,0} T\widetilde{f} = Tf_o
\end{aligned}$$

We used for the first equality that $M_{Tf_o,0} = M^*_{Tf_o,0} = M^2_{Tf_o,0}$. This proves that $\widetilde{f}$ belongs to the set $\mathcal{M}_{f_o}$.

Let us now prove that $\|\!|\widetilde{f}|\!\|_{\mathbf{w},p} \leq \|\!|f_o^\dagger|\!\|_{\mathbf{w},p}$. Because of the weak convergence of the $f_n^\star$ to $\widetilde{f}$, for all λ, the nonnegative sequence $\{w_\lambda|f_{n\lambda}^\star|\}_n$ converges to $w_\lambda|\widetilde{f}_\lambda|$. One can then use Fatou's lemma to obtain :

$$\|\!|\widetilde{f}|\!\|_{\mathbf{w},p}^p = \sum_\lambda \lim_{n\to\infty} \{w_\lambda|f_{n\lambda}^\star|\}_n \leq \lim_{n\to\infty} \sum_\lambda \{w_\lambda|f_{n\lambda}^\star|\}_n = \lim_{n\to\infty} \|\!|f_n^\star|\!\|_{\mathbf{w},p}^p$$

But we proved earlier that $\limsup_n \|\!|f_n^\star|\!\|_{\mathbf{w},p}^p \leq \|\!|f_o^\dagger|\!\|_{\mathbf{w},p}^p$. Therefore, we get :

$$\|\!|\widetilde{f}|\!\|_{\mathbf{w},p}^p \leq \lim_{n\to\infty} \|\!|f_n^\star|\!\|_{\mathbf{w},p}^p \leq \|\!|f_o^\dagger|\!\|_{\mathbf{w},p}^p \tag{2.32}$$

By definition, $f_o^\dagger$ is the unique minimizer of the $\|\!|.|\!\|_{\mathbf{w},p}$-norm in $\mathcal{M}_{f_o}$, so this implies that $\widetilde{f} = f_o^\dagger$.

The conclusion of this paragraph is that $f_o^\dagger$ is the only possible accumulation point of the sequence $f_n^\star$.

The sequence $\{f_n^\star\}_n$ converges weakly to $f_o^\dagger$:

We proved that the sequence $\{f_n^\star\}_n$ is uniformly bounded in the $\|.\|_{\mathcal{H}_1}$-norm and that it has a unique accumulation point : $f_o^\dagger$. This allows us to conclude that $f_n^\star$ converges weakly to $f_o^\dagger$.

The sequence $\{f_n^\star\}_n$ converges strongly to $f_o^\dagger$:

Replacing $\widetilde{f}$ by its value $f_o^\dagger$ in (2.32), we get : $\|\!|f_o^\dagger|\!\|_{\mathbf{w},p}^p \leq \lim_{n\to\infty} \|\!|f_n^\star|\!\|_{\mathbf{w},p}^p \leq \|\!|f_o^\dagger|\!\|_{\mathbf{w},p}^p$ which proves that the sequence $\{\|\!|f_n^\star|\!\|_{\mathbf{w},p}^p\}_n$ converges to $\|\!|f_o^\dagger|\!\|_{\mathbf{w},p}^p$. We shall see now that the two results we obtained so far :

$$f_n^\star \xrightarrow[n\to\infty]{w} f_o^\dagger \tag{2.33}$$

$$\|\!|f_n^\star|\!\|_{\mathbf{w},p} \xrightarrow[n\to\infty]{} \|\!|f_o^\dagger|\!\|_{\mathbf{w},p} \ , \tag{2.34}$$

imply the strong convergence of the sequence $\{f_n^\star\}_n$ to $f_o^\dagger$. (This argument closely follows [16].)

Let us prove that $\{\|f_n^\star\|_{\mathcal{H}_1}\}_n$ converges to $\|f_o^\dagger\|_{\mathcal{H}_1}$. We have :

$$\Big|\|f_n^\star\|_{\mathcal{H}_1}^2 - \|f_o^\dagger\|_{\mathcal{H}_1}^2\Big| = \Big|\sum_\lambda \big(|f_{n\lambda}^\star|^2 - |f_{o\lambda}^\dagger|^2\big)\Big| \leq \sum_\lambda \Big|\,|f_{n\lambda}^\star|^2 - |f_{o\lambda}^\dagger|^2\,\Big| \tag{2.35}$$

Writing $x^2 = (x^p)^{\frac{2}{p}}$ and using the derivability of $x \to x^{\frac{2}{p}}$, one can bound the last term :

$$\Big|\,|f_{n\lambda}^\star|^2 - |f_{o\lambda}^\dagger|^2\Big| \leq \tfrac{2}{p}\ \max\{\big(|f_{n\lambda}^\star|^p\big)^{\frac{2}{p}-1}, \big(|f_{o\lambda}^\dagger|^p\big)^{\frac{2}{p}-1}\}\ \Big|\,|f_{n\lambda}^\star|^p - |f_{o\lambda}^\dagger|^p\,\Big| \tag{2.36}$$

$$\leq \tfrac{2}{p}\ \max\{|f_{n\lambda}^\star|^{2-p}, |f_{o\lambda}^\dagger|^{2-p}\}\ \Big|\,|f_{n\lambda}^\star|^p - |f_{o\lambda}^\dagger|^p\,\Big| \tag{2.37}$$

$$\leq \tfrac{2}{pc}\ \max\{|f_{n\lambda}^\star|^{2-p}, |f_{o\lambda}^\dagger|^{2-p}\}\ \Big|\, w_\lambda|f_{n\lambda}^\star|^p - w_\lambda|f_{o\lambda}^\dagger|^p\,\Big| \tag{2.38}$$

We saw in Eq. (2.45) that for any $f \in \mathcal{H}_1$ and $\lambda_o \in \Lambda$ $|f_{\lambda_o}| \leq c^{\frac{1}{p}} \|\!| f |\!\|_{\mathbf{w},p}$. Plugging this into Eq. (2.38) and summing over λ, we get :

$$\Big| \|f_n^\star\|_{\mathcal{H}_1}^2 - \|f_o^\dagger\|_{\mathcal{H}_1}^2 \Big| \leq \tfrac{2}{p} c^{-\frac{2}{p}} \ \max\{\|\!| f_n^\star |\!\|_{\mathbf{w},p}^{2-p}, \|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^{2-p}\} \sum_{\lambda \in \Lambda} \Big| \ w_\lambda |f_{n\lambda}^\star|^p - w_\lambda |f_{o\lambda}^\dagger|^p \ \Big| \quad (2.39)$$

Since $\{\|\!| f_n^\star |\!\|_{\mathbf{w},p}^p\}_n$ converges to $\|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^p$, for n large enough, $\max\{\|\!| f_n^\star |\!\|_{\mathbf{w},p}^{2-p}, \|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^{2-p}\}$ is bounded by $2\|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^{2-p}$. Defining $g_{c,p,f_o} = \frac{4}{p} c^{-\frac{2}{p}} \ \|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^{2-p}$, we get :

$$\begin{aligned} \Big| \|f_n^\star\|_{\mathcal{H}_1}^2 - \|f_o^\dagger\|_{\mathcal{H}_1}^2 \Big| \ &\leq \ g_{c,p,f_o} \textstyle\sum_{\lambda \in \Lambda} \Big| \ w_\lambda |f_{n\lambda}^\star|^p - w_\lambda |f_{o\lambda}^\dagger|^p \ \Big| \\ &\leq \ g_{c,p,f_o} \textstyle\sum_\lambda \Big(w_\lambda |f_{n\lambda}^\star|^p + w_\lambda |f_{o\lambda}^\dagger|^p - 2 w_\lambda \min\{|f_{n\lambda}^\star|, |f_{o\lambda}^\dagger|\}^p \Big) \\ &\leq \ g_{c,p,f_o} \Big(\|\!| f_n^\star |\!\|_{\mathbf{w},p}^p + \|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^p - 2 \textstyle\sum_\lambda w_\lambda \min\{|f_{n\lambda}^\star|, |f_{o\lambda}^\dagger|\}^p \Big) \end{aligned} \quad (2.40)$$

We already know that $\|\!| f_n^\star |\!\|_{\mathbf{w},p}^p \xrightarrow[n\to\infty]{} \|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^p$, we shall see now that the same holds for the last term in the previous inequality. Let us define the sequence $\{u_{n\lambda}\}_n$ for each λ by $u_{n\lambda} = w_\lambda \min\{|f_{n\lambda}^\star|, |f_{o\lambda}^\dagger|\}^p$. The weak convergence of the $f_n^\star$ to $f_o^\dagger$ implies that for each λ, $u_{n\lambda} \xrightarrow[n\to\infty]{} w_\lambda |f_{o\lambda}^\dagger|^p$. Moreover, for all n, $0 \leq u_{n\lambda} \leq w_\lambda |f_{o\lambda}^\dagger|^p$ and $\sum_\lambda w_\lambda |f_{o\lambda}^\dagger|^p = \|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^p < \infty$ so that by the dominated convergence theorem, $\lim\limits_{n\to\infty} \sum_\lambda u_{n\lambda} = \sum_\lambda \lim\limits_{n\to\infty} u_{n\lambda}$. Replacing the $u_{n\lambda}$ and their limits by their value, we obtain :

$$\lim_{n\to\infty} \sum_\lambda w_\lambda \min\{|f_{n\lambda}^\star|, |f_{o\lambda}^\dagger|\}^p = \|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^p.$$

Hence :

$$\Big(\|\!| f_n^\star |\!\|_{\mathbf{w},p}^p + \|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^p - 2 \sum_\lambda w_\lambda \min\{|f_{n\lambda}^\star|, |f_{o\lambda}^\dagger|\}^p \Big) \xrightarrow[n\to\infty]{} \|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^p + \|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^p - 2 \|\!| f_o^\dagger |\!\|_{\mathbf{w},p}^p = 0$$

so that by taking the limit as n goes to ∞ in Eq.(2.40), we can conclude that

$$\|f_n^\star\|_{\mathcal{H}_1} \xrightarrow[n\to\infty]{} \|f_o^\dagger\|_{\mathcal{H}_1}.$$

Using the identity $\|f_n^\star - f_o^\dagger\|_{\mathcal{H}_1} = \|f_n^\star\|_{\mathcal{H}_1} + \|f_o^\dagger\|_{\mathcal{H}_1} - 2\big\langle f_n^\star, \ f_o^\dagger \big\rangle$, this last result combined with the weak convergence of the $f_n^\star$ to $f_o^\dagger$ proves that the sequence $\{f_n^\star\}_n$ converges strongly in $\mathcal{H}_1$ to $f_o^\dagger$. ∎

Note that we did not need to assume that each g_n is in $\mathcal{H}_1^{T,\mathbf{w},p}$ to obtain stability. It could very well be that the functional $J_{\gamma_n,\mathbf{w},p,\tau_n;g_n}$ has several minimizers, in that case, depending on the choice of the starting element for the iterative algorithm 2.3.4, the element f_n^* might have different values. As a result, the sequence $\{f_n^*\}_n$ is not fixed by the parameters ϵ_n, γ_n, τ_n and g_n. However no matter which of these sequences f_n^* we consider, it will converge strongly to $f_o^\dagger$.

2.3.4 Example

We give here an example where the operator T is a multiplication and the iterative algorithms 2.2.4 and 2.3.4 are applied on the same noisy image g_ϵ, with the same parameter γ. $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$ is the orthonormal basis formed by the Haar wavelet. We chose $\tau = 2\sigma$, where σ is the standard deviation of the noise. The top row of Figure 2.3 shows the original image f (left) ; the function t corresponding to the operator T (second column) ; the image of f under T : $g = T(f) = f.t$ (third column) and the noisy observation g_ϵ (right). Below, the results of iterative algorithms 2.2.4 (on the left) and 2.3.4 (on the right) are displayed. Although the standard iterative algorithm (2.2.4) yields almost perfect reconstruction in this case, the adaptive projection algorithm does not recover the object f. Because of the projections, one wavelet coefficient in g_ϵ is not taken in account. This prevents the iterative algorithm to properly inverse the operator T.

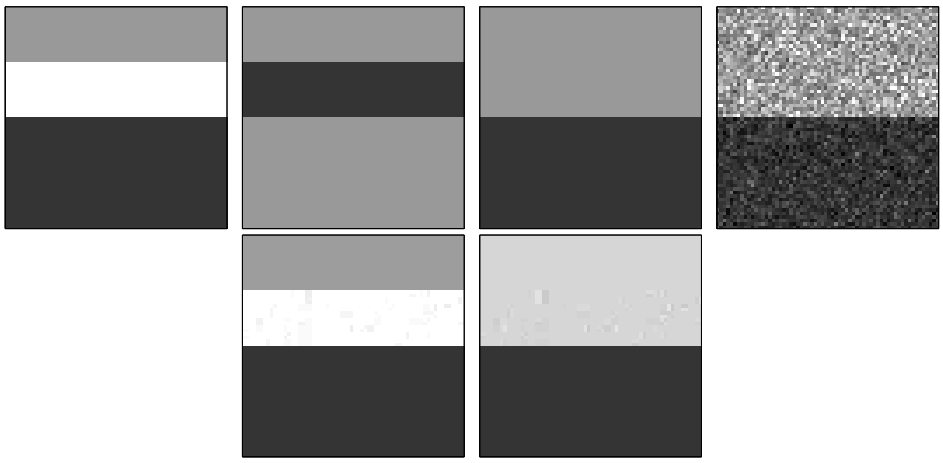

FIG. 2.3 – From left to right, top row : original f, multiplication operator t, image $g = t.f$, noisy observation of the image g_ϵ. Bottom row, left : reconstruction with the standard iterative algorithm ; right : reconstruction with adaptive projection.

2.4 Adaptive projections relaxed

Our discussion and example above showed that minimizing the adaptive projection functional may lead to an undesirable solution in some cases, depending on the operator T and the data. In this section, we introduce a slight relaxation of the adaptive projections that we will prove no longer suffers from this inconvenience.

2.4.1 Definition of the relaxed adaptive projections and of the corresponding iterative algorithm

Definition 2.4.1. *Given an orthonormal basis if $\mathcal{H}_2$, $\boldsymbol{\beta}=\{\beta_\lambda\}_{\lambda\in\Lambda}$, an element g in $\mathcal{H}_2$, a sequence of nonnegative thresholds $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$ and a scalar μ, $\mathrm{M}_{\mathrm{g},\tau,\mu}$ is the map from $\mathcal{H}_2$ into itself defined by :*

$$\forall h\in\mathcal{H}_2,\quad \mathrm{M}_{\mathrm{g},\tau,\mu}(h)=\sum_{\lambda\ s.t.\ |g_\lambda|>\tau_\lambda} h_\lambda\beta_\lambda+\mu\sum_{\lambda\ s.t.\ |g_\lambda|\leq\tau_\lambda} h_\lambda\beta_\lambda$$

Note that $\mathrm{M}_{\mathrm{g},\tau,\mu}$ is a bounded diagonal operator for any g, τ and μ. It is therefore a continuous linear operator. Depending on the parameters g and τ, either $\|\,\mathrm{M}_{\mathrm{g},\tau,\mu}\,\|=1$ or $\|\,\mathrm{M}_{\mathrm{g},\tau,\mu}\,\|=|\mu|$. In the following, we will restrict μ to the interval $(0,1]$ and therefore, we will always have $\|\,\mathrm{M}_{\mathrm{g},\tau,\mu}\,\|\leq 1$. Note that $\mathrm{M}_{\mathrm{g},\tau,0}$ is the adaptive projection defined before : $\mathrm{M}_{\mathrm{g},\tau,0}=\mathrm{M}_{\mathrm{g},\tau}$ and that, for any choice of g, τ and $\mu\neq 0$, $\mathrm{M}_{\mathrm{g},\tau,\mu}$ has a bounded linear inverse. The minimization problem now becomes :

Problem 2.4.2. *Given a sequence of strictly positive weights $\boldsymbol{w}=\{w_\lambda\}_{\lambda\in\Lambda}$, a sequence of nonnegative thresholds $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$, and scalars γ, μ and p with $\gamma>0$, $0<\mu\leq 1$ and $1\leq p\leq 2$, find :*

$$f^\star=\operatorname*{argmin}_{f\in\mathcal{H}_1}\ \mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau,\mu}(f)=\operatorname*{argmin}_{f\in\mathcal{H}_1}\ \|\,\mathrm{M}_{\mathrm{g},\tau,\mu}(Tf-g)\|_{\mathcal{H}_2}^2+\gamma|\!|\!|f|\!|\!|_{\mathbf{w},p}^p$$

where $|\!|\!|f|\!|\!|_{\mathbf{w},p}=\left[\sum_{\lambda\in\Lambda}w_\lambda|\left\langle f,\varphi_\lambda\right\rangle|^p\right]^{\frac{1}{p}}$ and

$$\mathrm{M}_{\mathrm{g},\tau,\mu}(h)=\sum_{\lambda\ s.t.\ |g_\lambda|>\tau_\lambda} h_\lambda\beta_\lambda+\mu\sum_{\lambda\ s.t.\ |g_\lambda|\leq\tau_\lambda} h_\lambda\beta_\lambda$$

For a fixed observation g and operator T, Problem 2.4.2 reduces to a particular instance of Problem 2.2.1, with the observation $g'=\mathrm{M}_{\mathrm{g},\tau,\mu}(g)$ and the operator $T'=\mathrm{M}_{\mathrm{g},\tau,\mu}\,T$. Therefore, the iterative algorithm that follows converges strongly to a minimizer of $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau,\mu}$ for any choice of the initial guess.

Algorithm 2.4.3.

$$\begin{cases} f^0 & \text{arbitrary} \\ f^n & = \mathbf{S}_{\gamma\mathbf{w},\mathrm{p}}\left(f^{n-1}+T^*\,\mathrm{M}_{\mathrm{g},\tau,\mu}^2\,g-T^*\,\mathrm{M}_{\mathrm{g},\tau,\mu}^2\,Tf^{n-1}\right),\ n\geq 1 \end{cases}$$

Theorem 2.4.4. *Let T be a bounded linear operator from $\mathcal{H}_1$ to $\mathcal{H}_2$, with norm strictly bounded by 1. Fix $p\in[1,2]$, $\mu\in(0,1]$, $\{\tau_\lambda\}_{\lambda\in\Lambda}$ a sequence of nonnegative numbers and let $\mathbf{S}_{\mathbf{w},\mathrm{p}}$ be the shrinkage operator defined by (2.7), where the sequence $\{w_\lambda\}_{\lambda\in\Lambda}$ is uniformly bounded below away from zero, i.e. there $\exists c>0$ s.t. $\forall\lambda\in\Lambda:\ w_\lambda\geq c$. Then the sequence of iterates*

$$f^n=\mathbf{S}_{\gamma\mathbf{w},\mathrm{p}}\left(f^{n-1}+T^*\,\mathrm{M}_{\mathrm{g},\tau,\mu}^2\,g-T^*\,\mathrm{M}_{\mathrm{g},\tau,\mu}^2\,Tf^{n-1}\right)\ ,\quad n=1,2,\dots\ ,$$

with f^0 arbitrarily chosen in $\mathcal{H}_1$, converges strongly to a minimizer of the functional

$$\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau,\mu}(f)=\|\,\mathrm{M}_{\mathrm{g},\tau,\mu}(Tf-g)\|_{\mathcal{H}_2}^2+\gamma|\!|\!|f|\!|\!|_{\mathbf{w},p}^p\ ,$$

where $\|\!|f|\!\|_{\mathbf{w},p}$ *denotes the norm* $\|\!|f|\!\|_{\mathbf{w},p} = \left[\sum_{\lambda\in\Lambda} w_\lambda |\langle f,\varphi_\lambda\rangle|^p\right]^{1/p}$, $1 \le p \le 2$ *and* $\mathrm{M}_{\mathrm{g},\tau,\mu}(h) = \sum_{\lambda \ s.t. \ |g_\lambda|>\tau_\lambda} h_\lambda\beta_\lambda + \mu \sum_{\lambda \ s.t. \ |g_\lambda|\le\tau_\lambda} h_\lambda\beta_\lambda$.

If the minimizer $f^\star$ *of* $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau,\mu}$ *is unique, (which is guaranteed e.g. by* $p > 1$ *or* $\ker(\mathrm{M}_{\mathrm{g},\tau,\mu} T) = \{0\}$*), then every sequence of iterates* f^n *converges strongly to* $f^\star$*, regardless of the choice of* f^0*.*

Démonstration. As we noticed before :

$$\begin{aligned}\mathbf{J}_{\gamma,\mathbf{w},p,\tau,\mu;T,g}(f) &= \|\,\mathrm{M}_{\mathrm{g},\tau,\mu}(Tf-g)\|^2_{\mathcal{H}_2} \ + \ \gamma\|\!|f|\!\|^p_{\mathbf{w},p} \\ &= \|(\mathrm{M}_{\mathrm{g},\tau,\mu}T)f - (\mathrm{M}_{\mathrm{g},\tau,\mu}\,g)\|^2_{\mathcal{H}_2} \ + \ \gamma\|\!|f|\!\|^p_{\mathbf{w},p} \\ &= \mathbf{J}_{\gamma,\mathbf{w},p,0,1;\ T',\ g'}(f) \quad \text{with} \quad T' = \mathrm{M}_{\mathrm{g},\tau,\mu}T, \ \ g' = \mathrm{M}_{\mathrm{g},\tau,\mu}\,g\end{aligned}$$

Noting that $\mathbf{J}_{\gamma,\mathbf{w},p,0,1;\ T',\ g'}(f)$ is exactly the functional defined in Problem 2.2.1, it is then sufficient to prove $\|\!|T'|\!\|$ is strictly smaller than 1 to prove the strong convergence of the iterative algorithm 2.4.3 via Theorem 2.2.5. But $\|\!|T'|\!\| = \|\!|\,\mathrm{M}_{\mathrm{g},\tau,\mu}T|\!\| \le \|\!|\,\mathrm{M}_{\mathrm{g},\tau,\mu}\,|\!\|.\|\!|T|\!\| \le \max\{1,|\mu|\}.\|\!|T|\!\|$. Since $0 < \mu \le 1$ then $\|\!|\,\mathrm{M}_{\mathrm{g},\tau}T|\!\| = 1$ and therefore $\|\!|T'|\!\| \le \|\!|T|\!\| < 1$. ∎

2.4.2 Stability

The difference between the relaxed adaptive projection functional $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau,\mu}$ and the original adaptive projection functional $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}$ is that we can now prove the desired stability result. We have, in analogy to Theorem 2.2.6 the following

Theorem 2.4.5. *Assume that* T *is a bounded operator from* $\mathcal{H}_1$ *to* $\mathcal{H}_2$ *with* $\|\!|T|\!\| < 1$ *and that the entries in the sequence* $\boldsymbol{w}=\{w_\lambda\}_{\lambda\in\Lambda}$ *are bounded below uniformly by a strictly positive number* c*.*

For any $g \in \mathcal{H}_2$ *and any* $\gamma > 0$, $0 < \mu \le 1$ *and nonnegative sequence* $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$*, define* $f^\star_{\gamma,\mathbf{w},p,\tau,\mu;\,g}$ *to be a minimizer of* $\mathbf{J}_{\gamma,\mathbf{w},p,\tau,\mu;\,g}(f)$*. If* $\gamma = \gamma(\epsilon)$, $\tau = \tau(\epsilon)$ *and* $\mu = \mu(\epsilon)$ *satisfy :*

1. $\lim_{\epsilon\to 0} \gamma(\epsilon) = 0$
2. $\lim_{\epsilon\to 0} \dfrac{\epsilon^2}{\gamma(\epsilon)} = 0$
3. $\forall\lambda\in\Lambda, \ \lim_{\epsilon\to 0} \tau_\lambda(\epsilon) = 0$
4. $\forall\lambda\in\Lambda, \ \exists\ \delta(\lambda) > 0, \quad s.t: \quad [\ \epsilon < \delta(\lambda) \Rightarrow \ \tau_\lambda(\epsilon) > \epsilon\ ]$
5. $\lim_{\epsilon\to 0} \mu(\epsilon) = \mu_o$*, with* $0 < \mu_o \le 1$

then for any f_o *such that there is a unique minimizer of the* $\|\!|\ |\!\|_{\mathbf{w},p}$*–norm in the set* $\mathcal{S}f_o = \{f : Tf = Tf_o\}$ *:*

$$\lim_{\epsilon\to 0}\left[\sup_{\|g-Tf_o\|_{\mathcal{H}_2}\le\epsilon} \|f^\star_{\gamma(\epsilon),\mathbf{w},p,\tau(\epsilon),\mu(\epsilon);\,g} - f_o^\dagger\|_{\mathcal{H}_1}\right] = 0\ ,$$

where $f_o^\dagger$ *is the unique element of minimum* $\|\!|\ |\!\|_{\mathbf{w},p}$*–norm in the set* $\mathcal{S}_{f_o}$*.*

Note that, if $\ker T = \{0\}$, then the set $\mathcal{S}_{f_o}$ reduces to f_o itself, so that the algorithm is regularizing for all element in $\mathcal{H}_1$. This ensures that when the noise level converges to 0, the sequence of estimates we obtain converges to the original object.

The proof of Theorem 2.4.5 is mostly analogous to (in fact a little easier than) the proof of Theorem 2.3.8. For the sake of completeness, we give the full details of the first two parts of the proof, indicating by

♠ ⇒

⇐ ♠

when the argument differs from before. Once we prove that $f_o^\dagger$ is the unique accumulation point of the sequence $\{f_n^\star\}_n$, the proof of weak and strong convergence are strictly identical and we shall not repeat them.

We start by a lemma that, similarly to Lemma 2.3.9, examines the convergence of the operators $\mathrm{M}_{g,\tau,\mu}$:

Lemma 2.4.6. *Suppose that $\tau = \tau(\epsilon)$ and $\mu = \mu(\epsilon)$ verify conditions 3, 4 and 5 of Theorem 2.4.5. Then the two following properties hold :*

1. *For any h in $\mathcal{H}_2$, $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}h$ converges weakly to $\mathrm{M}^2_{Tf,0,\mu_o}h$ as ϵ goes to 0.*
2. *If $h(\epsilon)$ converges weakly to h as ϵ goes to 0, then $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}h(\epsilon)$ converges weakly to $\mathrm{M}^2_{Tf,0,\mu_o}h$ as ϵ goes to 0.*

Proof of Lemma 2.4.6 : In the proof of Lemma 2.3.9, we have seen that under conditions imposed on $\tau(\epsilon)$ (conditions 3 and 4 of Theorem 2.4.5), the following happens :

- for λ s.t. $Tf_\lambda \neq 0$: $\lim_{\epsilon\to 0}\tau_\lambda(\epsilon) = 0$ hence there exists $\delta(\lambda, f)$ such that $\epsilon < \delta(\lambda, f)$ implies $\tau_\lambda(\epsilon) < |Tf_\lambda| - \epsilon$. It follows that : $|g(\epsilon, f)_\lambda| > \tau_\lambda(\epsilon)$.
- for λ s.t. $Tf_\lambda = 0$: $\epsilon < \delta(\lambda)$ implies $\tau_\lambda(\epsilon) > \epsilon$. It follows that if $\epsilon < \delta(\lambda)$, then $|g(\epsilon, f)_\lambda| > \tau_\lambda(\epsilon)$.

So that in the first case : $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}(\beta_\lambda) = \beta_\lambda$ for any $g(\epsilon, f)$ and any $\epsilon < \delta(\lambda, f)$; and in the second case : $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}(\beta_\lambda) = \mu(\epsilon)^2\beta_\lambda$ for any $g(\epsilon, f)$ and any $\epsilon < \delta(\lambda)$. Since $\mu(\epsilon)$ converges to some μ_o by assumption (condition 5 of Theorem 2.4.5), it follows that $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}h$ converges to $M^2_{Tf_o,0,\mu_o}h$ as (ϵ) goes to 0. This proves the first part of Lemma 2.4.6.

To prove the second part of Lemma 2.4.6, we use again the splitting trick we used in 2.3.9.(3) :

$$\left|\left[M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}h(\epsilon) - \mathrm{M}^2_{Tf,0,\mu_o}h\ \right]_\lambda\right| \tag{2.41}$$

$$= \left|\left[M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}(h(\epsilon) - h) + (M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)} - \mathrm{M}^2_{Tf,0,\mu_o})h\ \right]_\lambda\right| \tag{2.42}$$

$$= \left|\left[M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}(h(\epsilon) - h)\ \right]_\lambda\right| + \left|\left[(M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)} - \mathrm{M}^2_{Tf,0,\mu_o})h\ \right]_\lambda\right| \tag{2.43}$$

And the same argument as we used in Lemma 2.3.9.(3) allows to conclude. ∎

Note, that we did not need to prove this lemma that $0 < \mu_o \leq 1$. Now that the weak convergence of $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}$ is established, we proceed to the

Proof of Theorem 2.4.5 : Let us consider f_o in $\mathcal{H}_1$, that verifies that $\mathcal{S}_{f_o}$ has a unique minimizer $\|\!|\!|\ |\!|\!|_{\mathbf{w},p}$–norm. We note this minimizer $f_o^\dagger$. We fix the following sequences : $\{\epsilon_n\}_n$ such that $\epsilon_n \underset{n\to\infty}{\longrightarrow} 0$, $\{g_n\}_n$ such that $\forall n,\ \|g_n - Tf_o\|_{\mathcal{H}_2} \leq \epsilon_n$, and $\{\gamma_n\}_n \overset{def}{=} \{\gamma(\epsilon_n)\}_n$, $\{\mu_n\}_n \overset{def}{=} \{\mu(\epsilon_n)\}_n$ and $\{\tau_n\}_n \overset{def}{=} \{\tau(\epsilon_n)\}_n$ that verify conditions 1 to 5 in Theorem 2.4.5. For every n, we choose a minimizer $f_n^\star \overset{def}{=} f^\star_{\gamma_n,\mathbf{w},p,\tau_n,\mu_n;\, g_n}$ of the functional $J_n(f) \overset{def}{=} J_{\gamma_n,\mathbf{w},p,\tau_n,\mu_n;\, g_n}(f) = \|\, \mathrm{M}_{g_n,\tau_n,\mu_n}(Tf - g_n)\|^2_{\mathcal{H}_2} + \gamma_n |\!|\!| f |\!|\!|^p_{\mathbf{w},p}$. We want to prove that for any such choice of the ϵ_n, g_n, γ_n, μ_n, τ_n and $f_n^\star$, the sequence $f_n^\star$ converges strongly in $\mathcal{H}_1$ to $f_o^\dagger$. We will also note $M_n \overset{def}{=} \mathrm{M}_{g_n,\tau_n,\mu_n}$.

The sequences $\{|\!|\!| f_n^\star |\!|\!|_{\mathbf{w},p}\}_n$ and $\{\|f_n^\star\|_{\mathcal{H}_1}\}_n$ are uniformly bounded :

By definition of J_n, $\forall n$:

$$\begin{array}{rrcl} & |\!|\!| f_n^\star |\!|\!|^p_{\mathbf{w},p} & \leq & \frac{1}{\gamma_n} J_n(f_n^\star) \\ \text{so that} & |\!|\!| f_n^\star |\!|\!|^p_{\mathbf{w},p} & \leq & \frac{1}{\gamma_n} J_n(f_o^\dagger) \quad \text{since } f_n^\star \text{ minimizes } J_n. \end{array}$$

But : $\spadesuit \Rightarrow$

$$\begin{array}{rcll} J_n(f_o^\dagger) & = & \|M_n(Tf_o^\dagger - g_n)\|^2_{\mathcal{H}_2} + \gamma_n |\!|\!| f_o^\dagger |\!|\!|^p_{\mathbf{w},p} & \\ & = & \|M_n(Tf_o - g_n)\|^2_{\mathcal{H}_2} + \gamma_n |\!|\!| f_o^\dagger |\!|\!|^p_{\mathbf{w},p} & \text{since } Tf_o^\dagger = Tf_o \\ & \leq & \|M_n\|^2 . \|(Tf_o - g_n)\|^2_{\mathcal{H}_2} + \gamma_n |\!|\!| f_o^\dagger |\!|\!|^p_{\mathbf{w},p} & \\ & \leq & \max\{1, |\mu_n|^2\} . \epsilon_n^2 + \gamma_n |\!|\!| f_o^\dagger |\!|\!|^p_{\mathbf{w},p} & \text{since } \|Tf_o - g_n\| \leq \epsilon_n \end{array}$$

Hence

$$\forall n,\ |\!|\!| f_n^\star |\!|\!|^p_{\mathbf{w},p} \leq \max\{1, |\mu_n|^2\} . \frac{\epsilon_n^2}{\gamma_n} + |\!|\!| f_o^\dagger |\!|\!|^p_{\mathbf{w},p}. \tag{2.44}$$

Since $\frac{\epsilon_n^2}{\gamma_n} \underset{n\to\infty}{\longrightarrow} 0$ and $\mu_n \underset{n\to\infty}{\longrightarrow} \mu_o \in (0, 1]$, this proves that $\{|\!|\!| f_n^\star |\!|\!|_{\mathbf{w},p}\}_n$ is uniformly bounded. $\Leftarrow \spadesuit$

Moreover, $\mathbf{w}$ is bounded below by $c > 0$ and $p \leq 2$, so the $\|.\|_{\mathcal{H}_1}$-norm is bounded above by $c^{-\frac{1}{p}} |\!|\!| . |\!|\!|_{\mathbf{w},p}$:

$$|f_\lambda| = \big(|f_\lambda|^p\big)^{\frac{1}{p}} \leq \big(\tfrac{w_\lambda}{c} |f_\lambda|^p\big)^{\frac{1}{p}} \leq \big(\sum_{\lambda\in\Lambda} \tfrac{w_\lambda}{c} |f_\lambda|^p\big)^{\frac{1}{p}} = c^{-\frac{1}{p}} |\!|\!| f |\!|\!|_{\mathbf{w},p} \tag{2.45}$$

so that :

$$\|f\|^2_{\mathcal{H}_1} = \sum_{\lambda\in\Lambda} |f_\lambda|^2 \leq \sum_{\lambda\in\Lambda} \tfrac{w_\lambda}{c} |f_\lambda|^p |f_\lambda|^{2-p} \leq \sum_{\lambda\in\Lambda} \tfrac{w_\lambda}{c} |f_\lambda|^p \big[c^{-\frac{1}{p}} |\!|\!| f |\!|\!|_{\mathbf{w},p}\big]^{2-p} \tag{2.46}$$

$$\|f\|^2_{\mathcal{H}_1} \leq \tfrac{1}{c} |\!|\!| f |\!|\!|^p_{\mathbf{w},p} \big[c^{-\frac{1}{p}} |\!|\!| f |\!|\!|_{\mathbf{w},p}\big]^{2-p} = c^{-\frac{2}{p}} |\!|\!| f |\!|\!|^2_{\mathbf{w},p} \tag{2.47}$$

Hence, the sequence $\{f_n^\star\}$ is also uniformly bounded in $\mathcal{H}_1$.

$f_o^\dagger$ is the unique accumulation point of the sequence $\{f_n^\star\}_n$:

Since it is uniformly bounded in $\mathcal{H}_1$, the sequence $\{f_n^\star\}_n$ has at least one weakly convergent subsequence $\{f_k^\star\}_k$. Let us denote its weak limit $\widetilde{f}$. We shall now prove that $\widetilde{f} = f_o^\dagger$.

Since $f_k^\star$ is a minimizer of J_k obtained through the iterative algorithm, 2.4.3, it verifies the fixed point equation : $f_k^\star = \mathbf{S}_{\gamma_k \mathbf{w},\mathrm{p}}\left(f_k^\star + T^* M_k^2 g_k - T^* M_k^2 T f_k^\star\right)$. We note $h_k = f_k^\star + T^* M_k^2 g_k - T^* M_k^2 T f_k^\star$, so that $f_k^\star = \mathbf{S}_{\gamma_k \mathbf{w},\mathrm{p}}(h_k)$. By definition of the weak limit, it follows that :

$$\begin{array}{rcll}
\forall \lambda,\ \widetilde{f}_\lambda & = & \lim\limits_{k\to\infty} S_{\gamma_k w_\lambda}\big((h_k)_\lambda\big) & \\
& = & \lim\limits_{k\to\infty} \left[(h_k)_\lambda\right] + \lim\limits_{k\to\infty} \left[S_{\gamma_k w_\lambda}\left((h_k)_\lambda\right) - (h_k)_\lambda\right] & \text{but } \lim\limits_{k\to\infty} \gamma_k w_\lambda = 0 \\
\text{So, } \forall \lambda,\ \widetilde{f}_\lambda & = & \lim\limits_{k\to\infty} \left[(h_k)_\lambda\right] & \text{since } \forall x,\ S_v(x) \xrightarrow[v\to 0]{} x \\
& = & \lim\limits_{k\to\infty} \left[(f_k^\star + T^* M_k^2 g_k - T^* M_k^2 T f_k^\star)_\lambda\right] & \\
& = & \widetilde{f}_\lambda + \lim\limits_{k\to\infty} \left[(T^* M_k^2 g_k - T^* M_k^2 T f_k^\star)_\lambda\right] & \text{since } (f_k^\star)_\lambda \xrightarrow[k\to\infty]{} \widetilde{f}_\lambda.
\end{array}$$

As a result : $\forall \lambda,\ \lim\limits_{k\to\infty} \left[(T^* M_k^2 g_k - T^* M_k^2 T f_k^\star)_\lambda\right] = 0$.

♠ ⇒

Since $\|g_k - T f_o\| \leq \epsilon_k$, then $\|T^* M_k^2 (g_k - T f_o)\|_{\mathcal{H}_2} \leq \|T^*\| \|M_k\|^2 \epsilon_k < \max\{1, |\mu_k|\}^2 . \epsilon_k$. Since μ_k converges to $\mu_o \in (0, 1]$, and ϵ_k to 0, this proves that for all λ :

$$\lim_{k\to\infty} \left[(T^* M_k^2 T f_o - T^* M_k^2 T f_k^\star)_\lambda\right] = 0. \tag{2.48}$$

From Lemma 2.4.6.(1), we know that the sequence $\{M_k^2(T f_o)\}_k$ converges weakly to $M^2_{T f_o, 0, \mu_o}(T f_o) = T f_o$. ⇐ ♠

Together with the continuity of T^*, this leads to :

$$T^* M_k^2 T f_k^\star \xrightarrow[k\to\infty]{w} T^* T f_o. \tag{2.49}$$

On the other hand, $f_k^\star$ converges weakly to $\widetilde{f}$. Using the continuity of T, we get $T f_k^\star \xrightarrow[k\to\infty]{w} T\widetilde{f}$.

♠ ⇒

Lemma 2.4.6.(2) allows then to conclude that $M_k^2 T f_k^\star \xrightarrow[k\to\infty]{w} M^2_{T f_o, 0, \mu_o} T\widetilde{f}$ ⇐ ♠

and it follows from the continuity of T^* that :

$$T^* M_k^2 T f_k^\star \xrightarrow[k\to\infty]{w} T^* M^2_{T f_o, 0, \mu_o} T\widetilde{f}. \tag{2.50}$$

Plugging this last result in Eq. (2.49), we obtain the equality :

$$T^* M^2_{T f_o, 0, \mu_o} T\widetilde{f} = T^* T f_o \tag{2.51}$$

Note that $M_{T f_o, 0, \mu_o}$ is a self adjoint and that $M^2_{T f_o, 0, \mu_o}(T f_o) = M_{T f_o, 0, \mu_o}(T f_o) = T f_o$. Therefore the previous equality reduces to : $T^* M^2_{T f_o, 0, \mu_o} T(\widetilde{f} - f_o) = 0$. Taking the scalar product with $\widetilde{f} - f_o$, we obtain :

$\spadesuit \Rightarrow$

$$\begin{aligned}\left\langle \widetilde{f} - f_o,\ T^* M^2_{Tf_o,0,\mu_o} T(\widetilde{f} - f_o) \right\rangle = 0 \\ &\Leftrightarrow \left\langle M_{Tf_o,0,\mu_o} T(\widetilde{f} - f_o),\ M_{Tf_o,0,\mu_o} T(\widetilde{f} - f_o) \right\rangle = 0 \\ &\Leftrightarrow \|M_{Tf_o,0,\mu} T(\widetilde{f} - f_o)\|^2_{\mathcal{H}_2} = 0 \\ &\Leftrightarrow M_{Tf_o,0,\mu_o} T(\widetilde{f} - f_o) = 0 \\ &\Leftrightarrow T(\widetilde{f} - f_o) = 0 \quad \text{since } M_{Tf_o,0,\mu_o} \text{ is invertible.} \\ &\Leftrightarrow T\widetilde{f} = Tf_o\end{aligned}$$

This proves that $\widetilde{f}$ belongs to the set $\mathcal{S}_{f_o}$. $\Leftarrow \spadesuit$

Let us now prove that $\|\!|\widetilde{f}|\!\|_{\mathbf{w},p} \leq \|\!|f_o^\dagger|\!\|_{\mathbf{w},p}$. Because of the weak convergence of the $f_n^\star$ to $\widetilde{f}$, for all λ, the nonnegative sequence $\{w_\lambda |f^\star_{n\lambda}|\}_n$ converges to $w_\lambda|\widetilde{f}_\lambda|$. One can then use Fatou's lemma to obtain :

$$\|\!|\widetilde{f}|\!\|^p_{\mathbf{w},p} = \sum_\lambda \lim_{n\to\infty} \{w_\lambda |f^\star_{n\lambda}|\}_n \leq \lim_{n\to\infty} \sum_\lambda \{w_\lambda |f^\star_{n\lambda}|\}_n = \lim_{n\to\infty} \|\!|f_n^\star|\!\|^p_{\mathbf{w},p}$$

$\spadesuit \Rightarrow$
But we proved earlier that $\|\!|f_n^\star|\!\|^p_{\mathbf{w},p} \leq \max\{1, |\mu_n|\}.\frac{\epsilon_n^2}{\gamma_n} + \|\!|f_o^\dagger|\!\|^p_{\mathbf{w},p}$. Therefore, since the $\lim_{n\to\infty} \mu_n = \mu_o \in (0,1]$ and $\lim_{n\to\infty} \frac{\epsilon_n^2}{\gamma_n} = 0$, we get :

$$\|\!|\widetilde{f}|\!\|^p_{\mathbf{w},p} \leq \lim_{n\to\infty} \|\!|f_n^\star|\!\|^p_{\mathbf{w},p} \leq \|\!|f_o^\dagger|\!\|^p_{\mathbf{w},p} \tag{2.52}$$

$\Leftarrow \spadesuit$
By definition, $f_o^\dagger$ is the unique minimizer of the $\|\!|.|\!\|_{\mathbf{w},p}$-norm in $\mathcal{S}_{f_o}$, so this implies that $\widetilde{f} = f_o^\dagger$.

The conclusion of this paragraph is that $f_o^\dagger$ is the only possible accumulation point of the sequence $f_n^\star$.

The sequence $\{f_n^\star\}_n$ converges weakly to $f_o^\dagger$:

[This is identical to the proof given for Theorem 2.3.8]

The sequence $\{f_n^\star\}_n$ converges strongly to $f_o^\dagger$:

[This is identical to the proof given for Theorem 2.3.8] ∎

2.4.3 Example

To illustrate how the relaxation of the adaptive projection works in practice, let us revisit the example given in subsection 2.3.4. We chose $\mu = .5$ and ran the relaxed iterative algorithm on the data we presented in Figure 2.3. Figure 2.4 shows the original object we are trying to estimate (top), together with the result of each method (bottom). As we noticed before, the introduction of adaptive projections in

the discrepancy term prevents the iterative algorithm 2.3.4 to reconstruct the object. The bottom left panel of Figure 2.4 shows the "perfect" reconstruction obtained with the standard iterative algorithm of section 2.2. One can see in the middle panel at the bottom of the figure that the reconstruction of section 2.3 using adaptive projections misses one variation. The ability to recover the signal perfectly is regained by using the relaxed algorithm of section 2.4, as shown in the bottom right panel of Figure 2.4.

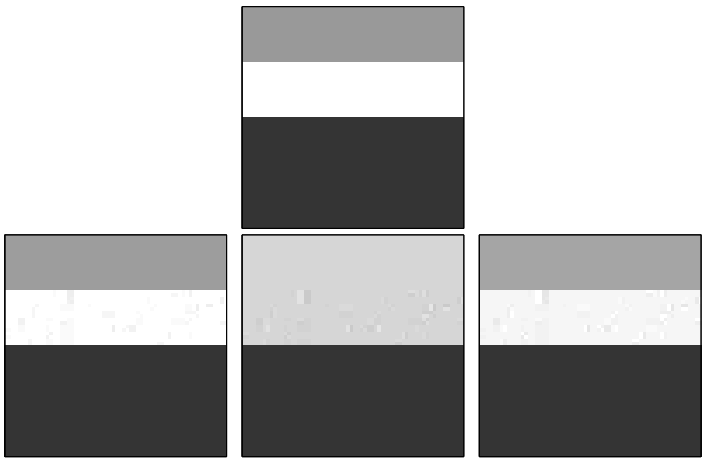

FIG. 2.4 – Example of Fig. 2.3 revisited. Top : original. Bottom, from left to right : reconstruction with the standard iterative algorithm, reconstruction with adaptive projections, reconstruction with relaxed adaptive projections.

2.5 Extension to multiple input/outputs

In this section, we discuss the generalization of the iterative algorithm to the case when one seeks M components $(f_1, f_2, \ldots, f_M)$ from L observations $(g_1, g_2, \ldots, g_L)$. We wish to minimize the functional defined in Eq.(2.1), choosing appropriate norms $|\!|\!|.|\!|\!|_{X_m}$ for each component f_m. As before, the norms $|\!|\!|.|\!|\!|_{X_m}$ are l_p-norms of decomposition coefficients. In all generality, the components f_m (resp. the observations g_l) could belong to different spaces Hilbert $\mathcal{H}_m^i$ (resp. $\mathcal{H}_l^o$). This would be the case, for instance, if one were to use this algorithm to register multi-modal data where each component could have a different format. One then needs to consider M tight frames $\boldsymbol{\varphi}^m = \{\varphi_\lambda^m\}_{\lambda\in\Lambda}$ for $m = 1, .., M$. Even if the components belong to the same Hilbert space, there is no reason a priori why the most appropriate norms $|\!|\!|.|\!|\!|_{X_m}$ would use the same tight frame for all $m = 1, .., M$. Therefore, we will allow not only the exponent p and the weights w_λ to depend on m but also the decomposition frame φ_λ :

$$|\!|\!|.|\!|\!|_{X_m} = \left[\sum_{\lambda\in\Lambda} w_\lambda^m |\left\langle\, .\,, \varphi_\lambda^m \right\rangle|^{p_m}\right]^{\frac{1}{p_m}}. \tag{2.53}$$

Note that we could introduce some modifications in the discrepancy terms as well, to tune these to the characteristics of each observation g_l, for $l = 1, .., L$. For example, one could use the (relaxed) adaptive projections $\mathrm{M}_{\mathrm{g},\tau,\mu}$. As is the case for $M = L = 1$, this amounts to modifying the operators and the observations g_l accordingly. Since we described in detail how these changes affect the iterative algorithm for $M = L = 1$, we shall focus here on the changes due to the presence of multiple observations and multiple components with specific $|\!|\!|.|\!|\!|_{X_m}$-norms. Subsection 2.5.1 describes the theoretical generalization of the iterative algorithm to the multiple components/multiple observations case and Subsection 2.5.2 the application to our astrophysical problem.

2.5.1 Generalization of the iterative algorithm

Let us first state the most general problem. Assuming we are given observations g_l that belong to different Hilbert spaces $\mathcal{H}_l^o$, we wish to estimate the objects f_m in Hilbert spaces $\mathcal{H}_m^i$ that produced them, knowing that the contribution of object f_m to observation g_l is $T_{m,l} f_m$ where the $T_{m,l} : \mathcal{H}_m^i \to \mathcal{H}_l^o$ are bounded linear operators. We estimate the objects f_m by solving the problem :

Problem 2.5.1. *Given scalars $\{\gamma_m\}_{m=1,..,M}$, $\{\rho_l\}_{l=1,..,L}$ and exponents $\{p_m\}_{m=1,..,M}$ with $\gamma_m > 0$, $\rho_l > 0$ and $1 \leq p_m \leq 2$, given in addition a tight frame $\boldsymbol{\varphi}^m = \{\varphi_\lambda^m\}_{\lambda\in\Lambda}$ and a sequence of positive weights $\boldsymbol{w}^m = \{w_\lambda^m\}_{\lambda\in\Lambda}$ for each Hilbert space $\mathcal{H}_m^i$, for $m = 1, .., M$, find :*

$$\underset{f_m\in\mathcal{H}_m^i}{\operatorname{argmin}}\ J(f_1, f_2, \ldots, f_M) = \sum_{l=1}^{L} \rho_l \Big\| \sum_{m=1}^{M} T_{m,l} f_m - g_l \Big\|_{\mathcal{H}_l^o}^2 + \sum_{m=1}^{M} \gamma_m |\!|\!| f_m |\!|\!|_{X_m}^{p_m} \ ;$$

where $|\!|\!| f |\!|\!|_{X_m} = \left[\sum_{\lambda\in\Lambda} w_\lambda |\left\langle f, \varphi_\lambda^m \right\rangle|^{p_m}\right]^{\frac{1}{p_m}}$.

Let us first explain the generalization of the iterative algorithm 2.2.4 needed to solve Problem 2.5.1, in the case where the p_m are equal.

Constant exponent : $p_m = p,\ \forall m$

When the exponents p_m are all the same, one can see Problem 2.5.1 as an instance of Problem 2.2.1 by recasting the Problem in higher dimension. This is done by building a unique observation space : $\overline{\mathcal{H}}^o = \mathcal{H}_1^o \times \mathcal{H}_2^o \times \cdots \times \mathcal{H}_M^o$ and a unique object space : $\overline{\mathcal{H}}^i = \mathcal{H}_1^i \times \mathcal{H}_2^i \times \cdots \times \mathcal{H}_L^i$. The standard euclidean norm :

$$\left\|\bar{f}\right\|_{\overline{\mathcal{H}}^i} = \left[\ \sum_{m=1}^{M} \left\|f_m\right\|_{\mathcal{H}_m^i}^2\right]^{\frac{1}{2}} \quad \text{for} \quad \bar{f} = (f_1, f_2, \ldots, f_M) \in\ \overline{\mathcal{H}}^i \tag{2.54}$$

defines $\overline{\mathcal{H}}^i$ as a Hilbert space. We define a particular norm on the Hilbert space $\overline{\mathcal{H}}^o$:

$$\left\|\bar{g}\right\|_{\overline{\mathcal{H}}^o} = \left[\ \sum_{l=1}^{L}\ \rho_l\ \left\|g_l\right\|_{\mathcal{H}_l^o}^2\right]^{\frac{1}{2}} \quad \text{for} \quad \bar{g} = (g_1, g_2, \ldots, g_M) \in\ \overline{\mathcal{H}}^o \tag{2.55}$$

Define the embedding operators $P_m : \mathcal{H}_m^i \rightarrow \overline{\mathcal{H}}^i$ by $P_m(f) = (0, \ldots, 0, \overset{m}{\overset{\downarrow}{f}}, 0, \ldots, 0)$. Since the family $\Phi = \{P_m(\varphi_\lambda^m)\}_{m=1,..,M,\, \lambda\in\Lambda}$ is a tight frame of $\overline{\mathcal{H}}^i$, one can also define a $|\!|\!|.|\!|\!|_X$-norm on the object space $\overline{\mathcal{H}}^i$ by :

$$|\!|\!|\bar{f}|\!|\!|_{\overline{\mathbf{w}},p} = \Big[\sum_{\substack{m=1,..,M\\ \lambda\in\Lambda}} \gamma_m\ w_\lambda^m\ \big|\big\langle \bar{f}, P_m(\varphi_\lambda^m)\big\rangle_{\overline{\mathcal{H}}^i}\big|^p \Big]^{\frac{1}{p}} = \Big[\sum_{\substack{m=1,..,M\\ \lambda\in\Lambda}} \gamma_m\ w_\lambda^m\ \big|\langle\, f_m, \varphi_\lambda^m\rangle_{\mathcal{H}_m^i}\big|^p \Big]^{\frac{1}{p}} \tag{2.56}$$

where $\overline{\mathbf{w}} = \{\gamma_m\, w_\lambda^m\}_{m=1,..,M,\, \lambda\in\Lambda}$. Finally the operator $\overline{T} :\ \overline{\mathcal{H}}^i \rightarrow \overline{\mathcal{H}}^o$ is defined by :

$$\overline{T}(f_1, f_2, \ldots, f_M) = \Big(\sum_{m=1}^M T_{m,1} f_m,\ \sum_{m=1}^M T_{m,2} f_m,\ \ldots,\ \sum_{m=1}^M T_{m,L} f_M\Big) \tag{2.57}$$

With these definitions, Problem 2.5.1 reduces to Problem 2.2.1 since :

$$J(f_1, f_2, \ldots, f_M) = \sum_{l=1}^L \rho_l \Big\| \sum_{m=1}^M T_{m,l} f_m - g_l \Big\|_{\mathcal{H}_l^o}^2 + \sum_{m=1}^M \gamma_m |\!|\!| f_m |\!|\!|_{X_m}^p \tag{2.58}$$

$$J(f_1, f_2, \ldots, f_M) = \Big\|\ \overline{T}\bar{f} - \bar{g}\ \Big\|_{\overline{\mathcal{H}}^o}^2 + |\!|\!|\bar{f}|\!|\!|_{\overline{\mathbf{w}},p}^p \tag{2.59}$$

$$\text{with } |\!|\!|\bar{f}|\!|\!|_{\overline{\mathbf{w}},p} = \Big[\sum_n \overline{w}_n \big|\big\langle \bar{f}, \Phi_n\big\rangle_{\overline{\mathcal{H}}^i}\big|^p\Big]^{\frac{1}{p}}. \tag{2.60}$$

Here the indexes λ and m are combined into a single index n and $\overline{w}_n = \overline{w}_\lambda^m = \gamma_m w_\lambda^m$ and $\Phi_n = \Phi_\lambda^m = P_m(\varphi_\lambda^m)$.

As a result, the iterative algorithm 2.2.4 can be used on the vectorized quantities $(\bar{f},\, \bar{g},\, \overline{T},\, \ldots)$ to solve Problem 2.5.1 when the p_m are equal.

Full case : arbitrary p_m

In the case where the p_m depend on m, the vectorization trick does not allow to conclude right away because $|\!|\!|\bar{f}|\!|\!|_X$ can not be written as a single l_p norm. One needs to go back to the construction of the iterative algorithm 2.2.4 to see how to modify it. We note Φ_λ^m the element $P_m(\varphi_\lambda^m)$ of the frame Φ. As before, the functional :

$$J(\bar{f}) = \Big\|\ \overline{T}\bar{f} - \bar{g}\ \Big\|_{\overline{\mathcal{H}}^o}^2 + \sum_{m,\,\lambda} \overline{w}_\lambda^m \Big|\big\langle \bar{f}, \Phi_\lambda^m\big\rangle_{\overline{\mathcal{H}}^i}\Big|^{p_m} \tag{2.61}$$

is approximated by the surrogate functional :

$$J^{\bar{a}}(\bar{f}) = \Big\|\overline{T}\bar{f} - \bar{g}\Big\|_{\overline{\mathcal{H}}^o}^2 - \Big\|\overline{T}\bar{f} - \overline{T}\bar{a}\Big\|_{\overline{\mathcal{H}}^o}^2 + C\Big\|\bar{f} - \bar{a}\Big\|_{\overline{\mathcal{H}}^i}^2 + \sum_{m,\,\lambda} \overline{w}_\lambda^m \big|\big\langle \bar{f}, \Phi_\lambda^m\big\rangle_{\overline{\mathcal{H}}^i}\big|^{p_m} \tag{2.62}$$

for $C > |\!|\!|\overline{T}^*\overline{T}|\!|\!|$. The surrogate functional is again strictly convex and the equations decouple for each pair (m, λ). The minimizer $\bar{f}^{\star^{\bar{a}}}$ is again defined applying the operator

$S_{w,p}$ for each component :

$$\left\langle \bar{f}^{\star^{\bar{a}}}, \Phi_\lambda^m \right\rangle_{\overline{\mathcal{H}}^i} = S_{\overline{w}_\lambda^m,\, p_m}\left(\left\langle C\bar{a} + \overline{T}^*\bar{g} - \overline{T}^*\overline{T}\bar{a}, \Phi_\lambda^m \right\rangle_{\overline{\mathcal{H}}^i}\right) \tag{2.63}$$

$$\left\langle \bar{f}^{\star^{\bar{a}}}, \Phi_\lambda^m \right\rangle_{\overline{\mathcal{H}}^i} = S_{\gamma_m . w_\lambda^m,\, p_m}\left(\left\langle C\bar{a} + \overline{T}^*\bar{g} - \overline{T}^*\overline{T}\bar{a}, \Phi_\lambda^m \right\rangle_{\overline{\mathcal{H}}^i}\right) \tag{2.64}$$

Defining the operator : $\overline{\mathbf{S}}_{\overline{w},\bar{p}} : \overline{\mathcal{H}}^i \to \overline{\mathcal{H}}^i$ by :

$$\overline{\mathbf{S}}_{\overline{w},\bar{p}}(\bar{f}) = \sum_{\substack{m=1,..,M \\ \lambda\in\Lambda}} S_{\overline{w}_\lambda^m,\, p_m}\left(\left\langle \bar{f}, \Phi_\lambda^m \right\rangle_{\overline{\mathcal{H}}^i}\right)\Phi_\lambda^m \ , \tag{2.65}$$

one gets :

$$\bar{f}^{\star^{\bar{a}}} = \overline{\mathbf{S}}_{\overline{w},\bar{p}}\left(C\bar{a} + \overline{T}^*\bar{g} - \overline{T}^*\overline{T}\bar{a}\right) \tag{2.66}$$

The only difference with what we saw in Section 2.2 is that now, the operator applied to each coordinate does not have the same value of p anymore. However the vectorized operator $\overline{\mathbf{S}}_{\overline{w},\bar{p}}$ (with multiple values for p) inherits the properties of the vectorized operator $\mathbf{S}_{\mathbf{w},\mathrm{p}}$ (with a single value p) that ensure the strong convergence of the iterative algorithm obtained by minimizing a sequence of surrogate functionals as before. (The mathematical definition 2.5.3 follows). That is to say that $\overline{\mathbf{S}}_{\overline{w},\bar{p}}$ is a non-expansive and asymptotically regular operator, it has at least one fixed point and verifies two technical lemmas (lemma 3.17 and lemma 3.18 in [16]). These properties are conserved because these is only a finite number of values p_m.

Hence when $\|\overline{T}^*\overline{T}\| < C$, an iterative algorithm that converges strongly to a solution of Problem 2.5.1 is :

Algorithm 2.5.2.

$$\begin{cases} \bar{f}^0 \ \in \overline{\mathcal{H}^i} \text{ arbitrary} \\ \bar{f}^n \ = \ \frac{1}{C}\ \overline{\mathbf{S}}_{\overline{w},\bar{p}}\left(C\bar{f}^{n-1} + \overline{T}^*\bar{g} - \overline{T}^*\overline{T}\bar{f}^{n-1}\right),\ n \geq 1 \end{cases}$$

Going back to the original observations g_l and operators $T_{m,l}$, the algorithm 2.5.2 in the original spaces $\mathcal{H}_m^i$ is :

Algorithm 2.5.3.

$$\begin{cases} f_m^{\mathbf{0}} \in \mathcal{H}_m^i, \quad \text{arbitrary},\ \forall\ m \in [\![1, M]\!] \\ \forall\ n \geq 1,\ \forall\ m \in [\![1, M]\!],\ \forall\ \lambda \in \Lambda : \\ \langle f_m^{\mathbf{n}}, \varphi_\lambda^m \rangle = \frac{1}{C} S_{\gamma_m w_\lambda^m,\, p_m}\left(< Cf_m^{\mathbf{n}-1} + \sum_{l=1,..,L} \rho_l T_{m,l}^* g_l - \sum_{\substack{l=1,..,L \\ r=1,..,M}} \rho_l T_{m,l}^* T_{r,l} f_r^{\mathbf{n}-1},\ \varphi_\lambda^m >\right)\varphi_\lambda^m \end{cases}$$

with $\|\overline{T}^*\overline{T}\| < C$

One can express a possible value for C in terms of upper bounds on the norms of the combinations of $T_{m,l}^* T_{r,l}$; we won't do this explicitly for this general case, but show in the next subsection how to do it for our particular application.

Remark. This approach is a generalization of the method developed in [20] for $M = 2$, with $p_1 = 1$ and $p_2 = 2$.

2.5.2 Application to astrophysical data

We present in this section the use of the multiple input/multiple output iterative algorithm 2.5.3 for our astrophysical problem. The objects g_l at hand are images of a portion of the sky, acquired at different wavelengths. The dominant components f_m in the observations are : the Cosmic Microwave background (f_1), the clusters of galaxies (f_2), infrared point sources (f_3) and the galactic dust (f_4). We are mostly interested in reconstructing accurately the clusters of galaxies. To do so it is necessary to consider the other signals, f_1, f_3 and f_4, because at the wavelength we consider they dominate the clusters' signal.

The observed images g_l all have the same resolution and size and we want to reconstruct images of the components with the same resolution and size as well. Hence, in this case, the Hilbert spaces $\mathcal{H}_m^i$ and $\mathcal{H}_l^o$ are the same. We have chosen to embed our input and output images in the Hilbert space $\mathcal{H} = L^2([0,1] \times [0,1])$ with the canonical norm.

Each image acquired on the telescope is a superposition of the different images we are trying to estimate that is blurred and contaminated by noise. The blurring occurs because the ideal impulse response of the instrument is not perfect. It is instead well modeled by the convolution with a function that depends on the observed wavelength. This function is called a "beam" in astronomy. Moreover, the contribution of each component depends on the wavelength of observation because of their different physical characteristics. As a result, the observed images g_l can be modeled as :

$$g_l = b_l \ * \ \Big[\sum_{m=1}^{M} a_{m,l} \ f_m \Big] + n_l \tag{2.67}$$

where $*$ denotes the two-dimensional convolution ; $a_{m,l}$ is a scalar ; b_l is the beam at wavelength l ; and n_l models the noise. Sources of noise here are instrumental noise and other components we overlooked because they are not dominant.

With this method, our estimates of the physical components f_1, f_2, ... are minimizers of the functional 2.5.1, computed via Algorithm 2.5.3. The operators $T_{m,l}$ combine the convolution by the beam and the frequency dependence of component m :

$$T_{m,l} : \begin{array}{ccc} \mathcal{H} & \rightarrow & \mathcal{H} \\ f & \mapsto & a_{m,l} \ b_l * \ f \end{array} \tag{2.68}$$

The beams b_l are typically square integrable functions and therefore the $T_{m,l}$ are bounded linear operators. The adjoint of $T_{m,l}$ is :

$$T_{m,l}^* : \begin{array}{ccc} \mathcal{H} & \rightarrow & \mathcal{H} \\ f & \mapsto & \bar{a}_{m,l} \ \widetilde{b_l} * \ f \end{array} \quad \text{where} \ \ \widetilde{b_l}(x,y) = \bar{b_l}(-x,-y) \tag{2.69}$$

Choice of the parameter C

The norm $\|\!|\overline{T}^* \overline{T}|\!\|$ can be bounded by noticing that :

$$\Big[\overline{T}^* \overline{T}(f_1, \ldots, f_M)\Big]_l(x) \ = \ \sum_{l=1}^{L} \sum_{r=1}^{M} \ \bar{a}_{m,l} \ a_{r,l} \ \rho_l \ (b_l * \widetilde{b_l} * f_r)(x) \tag{2.70}$$

Computing the Fourier transform of the previous equation, we obtain :

$$\left[\overline{T}^*\overline{T}(f_1,\ldots,f_M)\right]_l^\wedge(\xi) \;=\; \sum_{l=1}^{L}\sum_{r=1}^{M} \bar{a}_{m,l}\; a_{r,l}\; \rho_l\; \left|\widehat{b_l}\right|^2(\xi)\; \widehat{f_r}(\xi)\;, \tag{2.71}$$

or, writing it in a matrix form :

$$\begin{pmatrix} \left[\overline{T}^*\overline{T}(f_1,\ldots,f_M)\right]_1^\wedge(\xi) \\ \vdots \\ \left[\overline{T}^*\overline{T}(f_1,\ldots,f_M)\right]_L^\wedge(\xi) \end{pmatrix} = A \begin{pmatrix} \rho_1\; \left|\widehat{b_1}\right|^2(\xi) & & 0 \\ & \ddots & \\ 0 & & \rho_L\; \left|\widehat{b_L}\right|^2(\xi) \end{pmatrix} A^* \begin{pmatrix} \widehat{f_1}(\xi) \\ \vdots \\ \widehat{f_M}(\xi) \end{pmatrix}, \tag{2.72}$$

where A is the $M \times L$ matrix with entries $a_{m,l}$. Noting $\rho B(\xi)$ the $L \times L$ diagonal matrix with entries $\rho_1\; \left|\widehat{b_l}\right|^2(\xi)$, and fixing ξ, one gets :

$$\forall\; \xi,\; \sum_{m=1}^{M}\left|\left[\overline{T}^*\overline{T}(f_1,\ldots,f_M)\right]_m^\wedge(\xi)\right|^2 \leq \|\!|A\; \rho B(\xi)\; A^*|\!\| \sum_{m=1}^{M}\left|\widehat{f_m}(\xi)\right|^2 \tag{2.73}$$

Assuming the beams b_l are integrable so that $\sup_\xi \left|\widehat{b_l}\right|^2(\xi) < \infty$, one can bound the matrix norm :

$$\forall\; \xi,\; \|\!|A\; \rho B(\xi)\; A^*|\!\| \;\leq \|\!|A\; \left(\sup_\xi \rho B(\xi)\right)\; A^*|\!\| \leq \sup_{l,\xi}\left(\rho_l \left|\widehat{b_l}\right|^2(\xi)\right) \|\!|AA^*|\!\| \tag{2.74}$$

Eq. (2.73) can be rewritten :

$$\forall\; \xi,\; \sum_{m=1}^{M}\left|\left[\overline{T}^*\overline{T}(f_1,\ldots,f_M)\right]_m^\wedge(\xi)\right|^2 \leq \sup_{l,\xi}\left(\rho_l \left|\widehat{b_l}\right|^2(\xi)\right) \|\!|AA^*|\!\| \sum_{m=1}^{M}\left|\widehat{f_m}(\xi)\right|^2 \tag{2.75}$$

Integrating this last equation in ξ gives a bound on the norm $\|\!|\overline{T}^*\overline{T}|\!\|$:

$$\|\!|\overline{T}^*\overline{T}|\!\| \leq \sup_{l,\xi}\left(\rho_l \left|\widehat{b_l}\right|^2(\xi)\right) \|\!|AA^*|\!\| \tag{2.76}$$

For our astrophysical problem the beams are Gaussian so the integrability condition is verified. We used $C = 2\|\!|\overline{T}^*\overline{T}|\!\|$.

Choice of the norms

We are most interested is the clusters of galaxies map f_2. Clusters of galaxies are rare objects in the sky. They are very compact, typically a few arcminutes wide, with a peak of intensity in the center and filaments on the outskirts. Because of their compactness and rarity, the clusters of galaxies are well described by a few large wavelet coefficients. The l^1 norm on the wavelet coefficients (which is in fact

equivalent to the Besov B_1^1-norm), has proved to be a good a regularization norm for such signals [8, 10, 44]. Hence, that is what we use to constrain the object f_2 :

$$\|\!|.|\!\|_{X_2} = \sum_{\lambda \in \Lambda} |\langle\, .\,, \varphi_\lambda \rangle| \tag{2.77}$$

where $\boldsymbol{\varphi} = \{\varphi_\lambda\}_{\lambda \in \Lambda}$is a tight frame of complex wavelets. We describe in detail in Chapter 4 Section 4.2 the dual tree complex wavelet transform that is used here.

The Cosmic Microwave Background component, f_1, is a smooth and slowly varying signal. It spreads across the whole sky. Moreover its power spectrum $|\widehat{f_1}(\xi)|^2$ is well studied and can therefore be used to constrain the estimate of f_1. This can be done by adding weights w_λ to the l^2-norm in wavelet space :

$$\|\!|.|\!\|_{X_1} = \sum_{\lambda \in \Lambda} w_\lambda |\langle\, .\,, \varphi_\lambda \rangle|^2 \tag{2.78}$$

As is the case for Sobolev spaces, for which one chooses $w_\lambda = w_{j,k} = 2^{\zeta j p}$ for appropriate ζ, we use weights $w_\lambda = w_{j,k}$ that depend only on the scale j of the wavelet $\varphi_{j,k}$ (not on the location k). They are defined as follows :

$$w_\lambda = w_{j,k} = \frac{\int \left|\widehat{\varphi_{j,0}}(\xi)\right|^2 d\xi}{\int P_1(\xi) \left|\widehat{\varphi_{j,0}}(\xi)\right|^2 d\xi} \tag{2.79}$$

where $P_1(\xi)$ is a template of the power spectrum of the CMB studied by astrophysicists.

The Galaxy Dust is also a smooth and slowly varying signal that spreads across all sky. Its power spectrum is not as well studied as the CMB, so we investigated the relevance of different Sobolev type norms to constrain its smoothness. We obtain the best results by choosing $w_\lambda = w_{j,k} = 2^{3j}$, i.e. :

$$\|\!|.|\!\|_{X_4} = \sum_{\lambda=(j,k) \in \Lambda} 2^{3j} |\langle\, .\,, \varphi_\lambda \rangle|^2 \tag{2.80}$$

The last signal f_3 comes from really small objects that emit in the infrared spectrum, called infrared point sources. These point sources are rather rare. Since they are so small, they appear under the resolution of any image, so that the extent of a point source is smaller than one single pixel. For this signal it is then natural to stay in the pixel domain, requiring that the estimate is a sparse as possible :

$$\|\!|.|\!\|_{X_3} = \sum_{pixel} \left| f_3(pixel) \right| \tag{2.81}$$

Note that one would ideally want to use the l^0-norm : $\sum_{pixel} \delta_{|f_3(pixel)| \neq 0}$. However, the functional would then not be convex. So, we choose the exponent p to stay as sparse as possible while keeping the convexity, which is $p = 1$. (In fact, Donoho has shown, and used in several papers, that in many cases an l^1-constraint is a good proxy for an L^0-bound ; see e.g. [22, 23, 24].)

Choice of the regularizing parameters

The principal source of noise in the astrophysical data we consider is controlled by the time of exposure to the portion of sky imaged. Astrophysicists therefore customarily provide, as part of their data, not only the g_l but also an estimation of the noise level σ_l in the image acquired. When the f_m are close to the truth, the l^{th} discrepancy term $||\sum_m T_{m,l} f_m - g_l||^2$ should be of the order of σ_l^2. To give equal importance to each discrepancy term, we set $\rho_l = \frac{1}{\sigma_l^2}$.

Similarly, we chose the parameters γ_m so that the regularization terms $|||f_m|||_{X_m}$ have the same order magnitude as each other but also as the discrepancy terms. The estimation of the order of magnitude of $|||f_m|||_{X_m}$ is done numerically using simulations of each component.

Positivity constraints

The clusters' signal and the point sources' signal are positive. We introduce these constraints using the projection step described in 2.2.4 for these two components.

Chapitre 3

Statistical method

In this section, we present a method of separation of blurred mixtures of components based on a statistical description of each component to be estimated. This method is largely inspired by the work of J. Portilla et al. [47]. In that paper, the authors present a method for deblurring natural images that is based on a statistical description of the unknown elements constituting the observation, namely, the "true" image and the noise (the point spread function causing the blurring is supposed to be known). In that framework, the "true" image is viewed as a realization of a random process F, and the noise as a realization of another random process N. Consequently, the observation is a realization of a random process G that is a known function of the previous ones : $G = T(F, N)$. The description of the characteristics of the two random processes F and N induces a statistical model for the random process G. In return, given a particular observation i.e. a particular instance of G, this model gives information about the plausible instances of F and N that produced it. Using this information, one can define a notion of best estimate for the instances of F and N that produced the observation in hand, which is to say an estimate of the "true" image and of the noise given the observation we have. Several standard techniques exist to carry out these estimations ; one can use e.g. a "maximum a posteriori" approach, or a "maximum likelihood estimator", etc.... Here, as in [47], we shall use a Bayes least square estimate, i.e. we estimate the "true" image by computing the maximizer of the conditional expectation of the process F given the observation.

Given this framework for estimation, one is left with choosing a model for the processes F and N, so that the observation gives a plausible estimate for F (which is the estimate of the "true" image). The choices made in [47] are based on knowledge that has been acquired by studying natural images and their properties. In particular, they use wavelet expansions : going to wavelet space helps separating the noise from the "true" image, because the noise energy is spread out across wavelet coefficients whereas the wavelet transform of a natural image is typically concentrated in a few large coefficients. The wavelet transformation has another advantage : it has been observed that the distribution of wavelet coefficients of natural images is not Gaussian ; whereas the noise is typically well modeled by a Gaussian process. Moreover, the structure present in natural images causes their wavelet coefficients to behave in a more coherent manner than the noise's coefficients. For instance, the presence of

an edge is reflected by relatively large wavelet coefficients, through different scales, at the location of the edge.

In [47], the authors propose a method that takes advantage of the knowledge we just described. They chose a particular wavelet transform, the *steerable pyramid*, and modeled "neighborhoods" of wavelet coefficients by Gaussian Scale Mixtures (GSM). These neighborhoods are sets of wavelet coefficients associated with the same location and that behave in a coherent manner. Modeling the behavior of the wavelet coefficients in these neighborhoods jointly (instead of each singly) buys power for the estimation by taking advantage of the coherence present in the "true" image and absent in the noise. Moreover, the Gaussian Scale Mixture is a family of probability distributions that can capture the non-Gaussianity of a signal ; it has proved to be useful for modeling the distribution of wavelet coefficients in natural images [60]. Once this model is completely characterized, the authors of [47] compute the Bayes least square estimate of the "true" image ; the use of the GSM model makes this estimate easy to compute.

We have extended this method to the case of blurred mixtures of components in order to extract the clusters of galaxies from observed astrophysical data. Although our components are not natural images, part of the reasoning here still holds. In particular, the use of neighborhoods of wavelets coefficients becomes crucial. Not all our components deviate a lot from Gaussianity (indeed the CMB signal is Gaussian !), therefore, distinguishing the noise from such components solely on the basis of the marginal distributions can not be done. Consequently, the coherence of wavelets coefficients in the same neighborhood is essential to make this distinction. Moreover, some signals (e.g. the clusters of galaxies) are much less intense than others, causing the amplitude of their wavelet coefficients to be too small to be detected one by one. Taking advantage of their coherence becomes necessary to lower the intensity threshold for detection of these signals. Note that the (non–)Gaussianity of the different components has a physical meaning : for example, the deviation from Gaussianity of the CMB gives astrophysicists an indication on how to understand the Universe. As the cluster signal is itself highly non–Gaussian, a bad estimation of the cluster signal "pollutes" the estimated CMB signal, and thus the astrophysical conclusions. Therefore, careful treatment of the (non–)Gaussianity of these signals is necessary. Using the Gaussian Mixture Model allows us to do so in a simple and efficient manner since both Gaussian and non–Gaussian signals can be modeled with the same formalism.

In this chapter, we will present the theoretical aspects of this model illustrated by some examples. In the first section, we describe in detail the different constituents of the statistical model of the different signals present in the observations. In particular, we show how to define neighborhoods of wavelet coefficients, what are Gaussian Scale Mixture models and what is the resulting model for each component. The second section discusses the formal derivation of the Bayes least square estimate and its computation, leaving the problem of the estimation of the different parameters for section three. Finally, we describe in the last section of this chapter the application of this method to our astrophysical problem. As we go along, we shall give some examples to illustrate the theoretical aspects of this method ; however most examples are kept until in Chapter 5, where we juxtapose the results produced by this method

and by the functional method of Chapter 2, so that the reader can easily compare them.

3.1 Modelization of the signals

Let us now give more explicits details about the different constituents of the statistical model of the data.

3.1.1 Neighborhoods of wavelet coefficients

In natural images, although the wavelet transform has the property of decorrelating coefficients, there exists significant spatial dependencies in the transformed coefficients : wavelet coefficients centered at the same (or a close) location and scale behave coherently. This is a consequence of the geometrical properties of such images and of the spatial localization of wavelets. For example, a vertical edge separating two smooth regions yields a recognizable pattern in the wavelet transform : all wavelet coefficients are very small, except those corresponding to a wavelet oriented horizontally and whose support includes the edge. Not only will the horizontal wavelets centered at the edge yield quite large coefficients, but also the horizontal coefficients will decay or oscillate in a special manner with the distance to the edge and with the scale. (In fact, if such a simple vertical discontinuity was located at $\overline{n} = 2^j(k_o^1, k_o^2)$, $(k_1, k_2) \in \mathbb{Z}^2$, one could derive the exact values taken by the coefficients $\left\langle f, \varphi^{vert}_{j',\overline{n}'} \right\rangle$ for scales j' finer than j, centered at locations $n' = 2^{j'}(k', k_o^2), k' \in [\![k_o^1 - K, k_o^1 + K]\!]$. Here we denoted φ^{vert} the wavelet that is vertically oriented).

Similarly, for our astrophysical problem, the geometrical properties of the different components can be exploited. For example, clusters of galaxies are spatially localized structures with a high intensity peak at their center. Their size is of the order of a couple arcminutes. Hence, at scales j where the width of the wavelet φ_j is a couple of arcminutes or less, the amplitude of wavelet coefficients should exhibit rather sharp transitions from very low to very high amplitude at the locations of the clusters. Moreover, these transitions should be sparsely distributed since the clusters are rare. This would not happen for the CMB signal (resp. the galaxy dust) for which the variations are much smoother and the typical scale of variations is more than 10 (resp. 50) times bigger. The point sources on the other hand are much less extended than the clusters and the noise is spread over scale and space. Hence the local behavior of the wavelet coefficients is particular to each component.

Different approaches have been proposed to take in account the spatial coherence of wavelet coefficients in order to improve image processing. The zerotree method for compression [55] and later the hidden Markov model based on wavelet trees for image denoising [15, 50] both incorporate the spatial dependencies as prior knowledge on the wavelet tree structure. Other methods are based on local models of the coefficients that are used either to compute parameters for the denoising [54] or as statistical prior for estimation of the signal [41, 47]. Most of these methods [55, 15, 50, 54] consider only the depencies between a wavelet coefficient and its parent (i.e. the coefficient centered

at the same location but at the next coarser scale). In our problem, the presence of the blurring will induce depencies on the wavelet coefficients within a scale as well. So we will use, as in [47, 41], more extended neighborhoods. We consider the neighborhood of a coefficient $f_{j,q,\overline{n}} = \left\langle f, \varphi^q_{j,\overline{n}} \right\rangle$ to be the set that contains the coefficient itself and its parent, $f_{j-1,q,\overline{n}}$, as well coefficients at the same scale j and orientation q, centered at positions $\overline{n}'$, where $\overline{n}'$ belongs to a $K-$ring of $\overline{n}$. Using the notation $\mathbf{f}_{j,q,\overline{n},K}$ for the neighborhood of the coefficient $f_{j,q,\overline{n}}$, this amounts to :

$$\mathbf{f}_{j,q,\overline{n},K} = \{f_{j-1,q,\overline{n}}\} \cup \{f_{j,q,\overline{n}'},\ \overline{n}' = \overline{n} + (i,j),\ (i,j) \in [\![-K,K]\!]^2\} \tag{3.1}$$

We note $\mathcal{V}_{j,q,\overline{n},K}$ the set of indexes of wavelet coefficients in the neighborhood $\mathbf{f}_{j,q,\overline{n},K}$:

$$\mathcal{V}_{j,q,\overline{n},K} = \{\ (j-1,q,\overline{n})\ \} \cup \{\ (j,q,\overline{n}'),\ \overline{n}' = \overline{n} + (i,j),\ (i,j) \in [\![-K,K]\!]^2\} \tag{3.2}$$

so that $\mathbf{f}_{j,q,\overline{n},K} = \{f_i\}_{i \in \mathcal{V}_{j,q,\overline{n},K}}$. Note that for $K = 0$, this reduces to the wavelet coefficient and its parent. For our application, $K = 1$ is typically sufficient to model the statistical dependences of the wavelet coefficients of the different components. Taking the blurring into account, we will extend the size of the neighborhood up to $K = 3$ to obtain a good estimation from the observations. For the sake of conciseness in the notation, we shall drop the index K indicating the size of the ring (and sometimes even the wavelet index $j, q, \overline{n}$) where not necessary, denoting the neighborhood $\mathbf{f}_{j,q,\overline{n},K}$ by $\mathbf{f}_{j,q,\overline{n}}$ (or even $\mathbf{f}$). Furthermore, the neighborhoods are ordered so that we describe them as vectors.

In [55, 15, 50, 41], the behavior of a single wavelet coefficient is described by a two-state model : a wavelet coefficient is either significant or not. The marginal distribution of a coefficient is a mixture of two centered Gaussians. One of them has small variance, this accounts for the high number of very small (i.e. non-significant) coefficients. The second Gaussian has a large variance, this accounts for the existence of large (i.e. significant) coefficients, giving more weight to the tail of the distribution than a single Gaussian would normally have. Because we want to model several components, we would like our model to offer the possibility of making a finer description of the behavior of wavelet coefficients. To do so, we use the Gaussian Scale Mixture model (GSM), also used in [47]. This model is more flexible than the two-state mixture of Gaussian model, allowing to fit a wide variety of marginal distributions.

3.1.2 Gaussian scale mixtures

Model

We model each neighborhood vector $\mathbf{f}$ as a Gaussian scale mixture. That is to say : the probability distribution of the vector $\mathbf{f}$ is the distribution of a product of two random variables, $\sqrt{z}$ and $\mathbf{u}$:

$$\mathbf{f} \overset{dist.}{=} \sqrt{z}\,\mathbf{u} \tag{3.3}$$

$\mathbf{u}$ is a centered Gaussian vector and z is a scalar random variable that takes only non-negative values. The random variable z, whose distribution we describe later, is

called the *multiplier* and is independent of $\mathbf{u}$. We shall always normalize z so that its expectation is one : $E\{z\} = 1$. It follows that the covariance matrix of the Gaussian vector $\mathbf{u}$ is exactly the covariance matrix of the neighborhood vector $\mathbf{f}$:

$$\begin{aligned}
\mathbf{Cov}(\mathbf{f}_i, \mathbf{f}_j) &= E\{\mathbf{f}_i\mathbf{f}_j\} - E\{\mathbf{f}_i\}E\{\mathbf{f}_j\} \\
&= E\{(\sqrt{z}\,\mathbf{u}_i)(\sqrt{z}\,\mathbf{u}_j)\} - E\{\sqrt{z}\,\mathbf{u}_i\}E\{\sqrt{z}\,\mathbf{u}_j\} \\
&= E\{z\}E\{\mathbf{u}_i\,\mathbf{u}_j\} - E\{\sqrt{z}\}^2 E\{\mathbf{u}_i\}E\{\mathbf{u}_j\} \qquad (\mathbf{u}\text{and z are independent}) \\
&= E\{\mathbf{u}_i\,\mathbf{u}_j\} \qquad (E\{z\} = 1,\ E\{\mathbf{u}_i\} = E\{\mathbf{u}_j\} = 0) \\
&= E\{\mathbf{u}_i\,\mathbf{u}_j\} - E\{\mathbf{u}_i\}E\{\mathbf{u}_j\} \qquad (\text{again } E\{\mathbf{u}_i\} = E\{\mathbf{u}_j\} = 0) \\
\mathbf{Cov}(\mathbf{f}_i, \mathbf{f}_j) &= \mathbf{Cov}(\mathbf{u}_i, \mathbf{u}_j)
\end{aligned}$$

The GSM model is then specified by two parameters : the probability distribution of the multiplier z, noted p_z, and the covariance matrix of $\mathbf{f}$, noted $\mathbf{C_f}$. Let us now describe how these two parameters affect the properties of the distribution of the vector $\mathbf{f}$.

Properties of the marginal distributions

From Eq. (3.3), one can see that the marginal distributions of the elements of $\mathbf{f}$ (i.e. the $p_{\mathbf{f}_i}$) may have different variances but all have the same shape. The variances of the marginal distributions are given by the diagonal of the matrix $\mathbf{C_f}$ whereas their common shape depends on the probability density of the multiplier, p_z.

If z is identically 1, then $\mathbf{f}_i = \mathbf{u}_i$, and therefore, the marginal distributions are Gaussian. By choosing another probability distribution for z, one can shape the marginal distributions of $\mathbf{f}$ to fit a wide range of distributions. In [1], Andrews and Mallows showed that for any scalar process x whose probability density function f_x is symmetric and verifies :

$$\left(-\frac{d}{dy}\right)^k f_x(y^{\frac{1}{2}}) \geq 0, \text{ for } y > 0,$$

one can find a multiplier z such that the corresponding Gaussian scale mixture has the same distribution as x. (This is actually also a necessary condition.) In Fig. 3.1, we plot the Gaussian probability density together with two examples of probability distributions that can be described by Gaussian scale mixtures : the Laplace distribution ($f_x = \frac{1}{2}e^{-|x|}$) and the logistic distribution ($f_x = \frac{e^{-x}}{(1+e^{-x})^2}$). The probability densities (f_x) are plotted on the left panel of the figure and their logarithm in base 10 ($\log_{10}(f_x)$) on the right panel. These probability densities have been scaled to have the same variance.

The graphs in Fig. 3.1 highlight two particular features of the marginal distributions that can be tuned using Gaussian scale mixtures. On the one hand, the behavior of the GSM at the origin can range from very smooth (like the Gaussian or the logistic distribution) to very "peaked" (like the exponential) This can be seen in the left panel of Fig. 3.1. On the other hand, a GSM distribution can have a fatter tail than the Gaussian distribution (see right panel of Fig. 3.1). Similarly, if a signal has

a very sparse wavelet expansion, most of its wavelet coefficients are small, therefore their probability density at the origin is rather peaky ; some coefficients, on the other hand, will be quite large, and therefore the tail of the probability distribution will be significantly fatter than the Gaussian [60]. These features typically model the non-Gaussian behavior of wavelet coefficients and we will exploit them later.

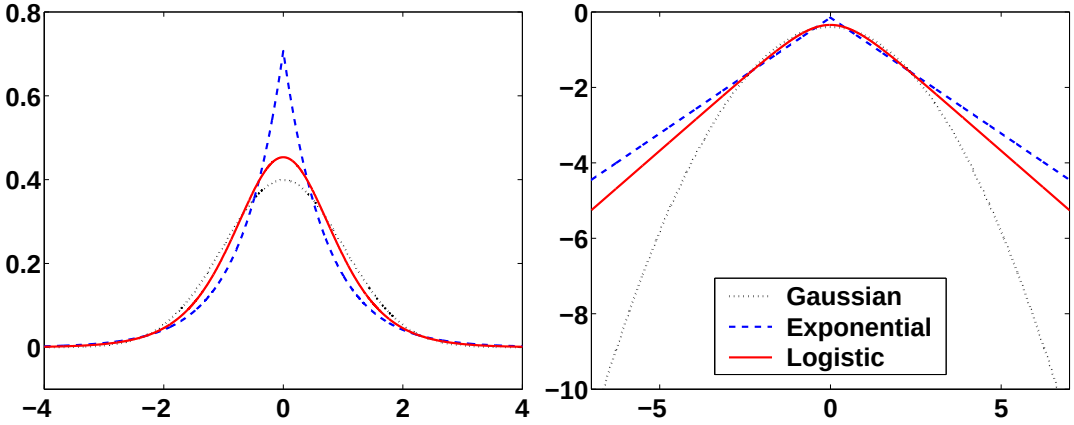

FIG. 3.1 – Probability density of several scalar Gaussian scale mixtures with the same variance : the Gaussian distribution, the exponential distribution and the logistic distribution. Left : the probability densities f_x, right : their logarithm $\log_{10}(f_x)$.

The example of the logistic distribution also shows that those two features can be tuned independently from each other : the logistic density is as smooth as the Gaussian density at the origin but still has heavy tails. At this point, we should mention that the two properties (behavior at the origin and at the infinities) are exactly the ones modeled by the two-state Gaussians. However, the probability density of the mixture of two Gaussians decays as the wider Gaussian, not enabling slower asymptotic decay ; it is smooth at the origin, and it is not differentiable. We hope that the flexibility of the Gaussian Scale Mixture will enable us to fit the experimental distribution of the wavelet coefficients of the clusters of galaxies more precisely than what we would obtain with a mixture of two Gaussians.

Properties of the conditional distributions

As we have just seen, the introduction of the multiplier z in the Gaussian scale mixture gives the possibility to fit a wide variety of marginal distributions. We shall now see how the multiplier also affects the conditional distributions in the GSM model. When the GSM is used for neighborhoods of wavelet coefficients, these conditional distributions, together with the covariance matrices $\mathbf{C_f}$ model the spatial dependencies between the coefficients. We have shown earlier that the covariance matrices of the vectors $\mathbf{f}$ and $\mathbf{u}$ are the same. Hence the "averaged" dependencies between two elements in $\mathbf{f}$ is captured in the model by the Gaussian vector $\mathbf{u}$. These depencies are however further tuned by the multiplier.

To illustrate this, let us consider several two-dimensional GSM that all have the identity matrix as their covariance matrix. The GSM model in two-dimensions is then : $(x_1, x_2) \overset{dist.}{=} (\sqrt{z}\ u_1, \sqrt{z}\ u_2)$. Here, x_1, x_2, u_1, u_2 and z are scalar random variables ; x_1, x_2, u_1 and u_2 are centered. The choice of the identity as a covariance matrix imposes that x_1, x_2, u_1 and u_2 have unit variance ; that x_1 and x_2 (resp. u_1 and u_2) are decorrelated ; and that the joint density of x_1 and x_2, $p_{(x_1,x_2)}$, is radial.

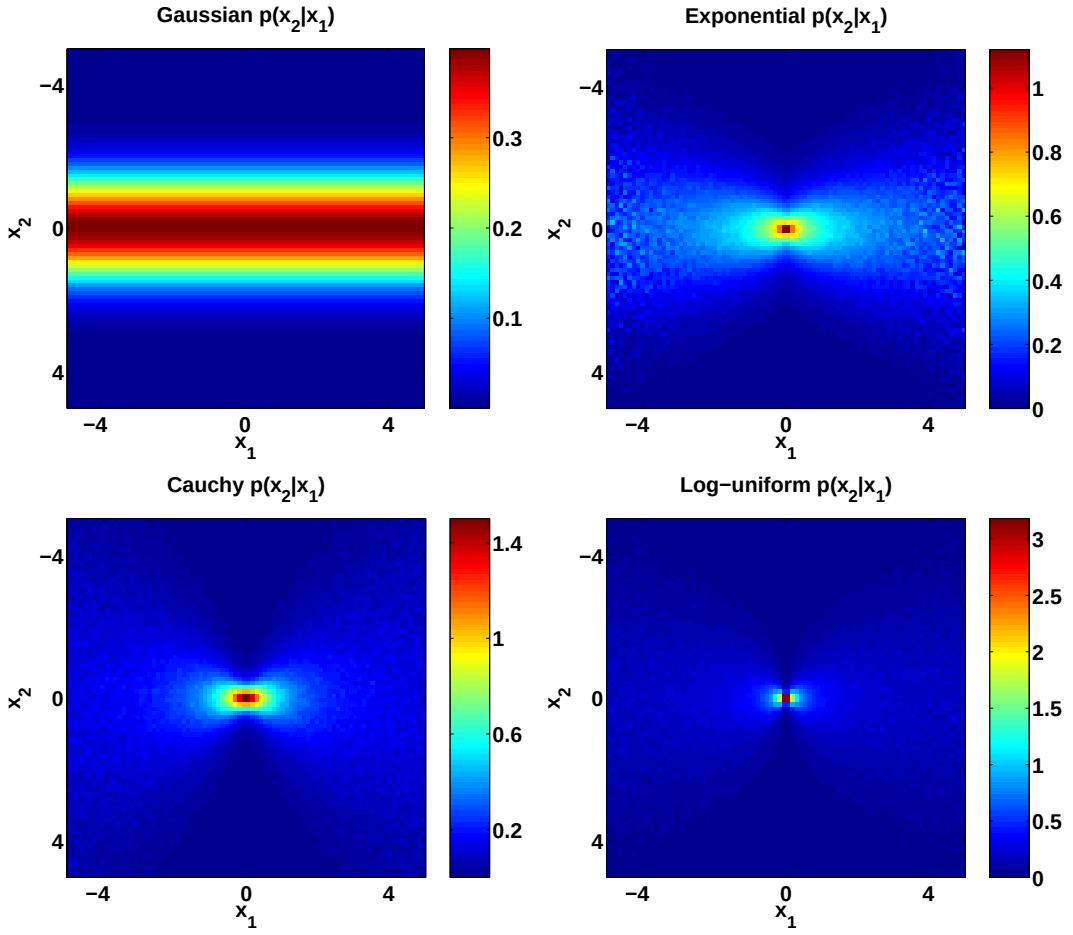

FIG. 3.2 – Conditional probability density $p(x_2|x_1)$ of several two-dimensional Gaussian scale mixtures with the same covariance matrix : the Identity matrix. Left to right, then up and down : the Gaussian distribution, the exponential distribution, the Cauchy distribution and the log-uniform distribution.

Note that, since u_1 and u_2 form a Gaussian vector and are decorrelated, they are independent whereas x_1 and x_2 are not independent, unless z is identically 1. Hence, although they have the same covariance matrix, $\mathbf{u}$ and $\mathbf{f}$ do not need to have the same conditional distributions. The presence of the multiplier z in the GSM allows to shape the conditional distribution of x_2 given x_1, $p(x_2|x_1)$, differently. In Fig. 3.2, the conditional distributions $p(x_2|x_1)$ are plotted for different GSM with the Identity as a covariance matrix. Each column of a plot represents the conditional probability

density of x_2 given x_1, for a fixed value of x_1. The top left panel shows the conditional probability in the case where the GSM reduces to a Gaussian vector ($z = 1$). Since x_1 and x_2 are independent in that case, the conditional probability $p(x_2|x_1)$ is the same for all values of x_1. In the other cases, the multiplier's distribution is not trivial and consequently, the conditional probability $p(x_2|x_1)$ depends on the value of x_1.

The non-Gaussian distributions displayed in Fig. 3.2 (top right and bottom left and right) exhibit a bow-tie shape that has been observed for neighboring wavelet coefficients in natural images [60]. The characteristics of a bow-tie shape distribution are : the conditional distribution $p(x_2|x_1)$ is concentrated around zero when the absolute value of x_1 is small, but much more spread out for larger values of x_1. For neighborhood of wavelet coefficients, this translates into : if the central coefficient is very small, its neighbors are typically very small as well ; if the central coefficient is very large, its neighbors can take a much larger set of values.

3.1.3 Resulting model for each component

As we stated in section 2.5.2, our astrophysical problem is to reconstruct several objects f^1, f^2, .., f^M from noisy and blurred observations of mixtures of them, g^1, g^2,.., g^L, determined by equation (2.67). (We use in this chapter superscripts for the indexes of the components and observations, since it makes the notation easier for the corresponding neighborhoods of wavelet coefficients). Our a priori model for each object f^m is that the statistical behavior of the neighborhoods of wavelet coefficients $\mathbf{f}^m_{j,q,\overline{n},K}$ can be described by a Gaussian scale mixture.

The physical properties of one component are identical in every direction and in every spatial location. Therefore, it seems that the modelization of a neighborhood $\mathbf{f}^m_{j,q,\overline{n},K}$ should depend only on m, K and the scale j, and not on $\overline{n}$ nor q, leading to $\mathbf{f}^m_{j,q,\overline{n},K} \overset{dist.}{=} \sqrt{z^m_{j,K}}\mathbf{u}^m_{j,K}$. However, we will need to keep the dependence in orientation q in the Gaussian vector $\mathbf{u}^m_j$. Indeed, a neighborhood $\mathbf{f}^m_{j,q,\overline{n},K}$ contains the parent coefficient, $f^m_{j-1,q,\overline{n},K}$, and a "square" neighborhood of coefficients at the same scale : $f^m_{j-1,q,\overline{n}',K}$, for $\overline{n}' = \overline{n} + (i,j)$, $(i,j) \in [\![-K,K]\!]^2$. Therefore the neighborhood $\mathbf{f}^m_{j,q,\overline{n},K}$ is not the rotated version of the neighborhood $\mathbf{f}^m_{j,0,\overline{n},K}$. Moreover, we will need to order the neighborhoods $\mathbf{f}^m_{j,q,\overline{n},K}$ into vectors with the same order regardless of the orientation. For example if $K = 1$, we will use the order :

$$\begin{aligned}\mathbf{f}_{j,q,0,1} \;=\; (&f_{j,q,(-1,-1)},\ f_{j,q,(-1,0)},\ f_{j,q,(-1,1)},\ f_{j,q,(0,-1)},\ f_{j,q,(0,0)},\ f_{j,q,(0,1)},\\ &f_{j,q,(1,-1)},\ f_{j,q,(1,0)},\ f_{j,q,(1,1)},\ f_{j-1,q,(0,0)}).\end{aligned}$$

The first two terms, $f_{j,q,(-1,-1)}$ and $f_{j,q,(-1,0)}$, always correspond to wavelets that are each other's shifts in the vertical direction and therefore their correlation depends on the orientation q of the wavelet. (Note that this problem would still arise with "circular" neighborhoods.) We are left with a model of the form : $\mathbf{f}^m_{j,q,\overline{n},K} \overset{dist.}{=} \sqrt{z^m_{j,K}}\ \mathbf{u}^m_{j,q,K}$.

The size K of the neighborhood we have to consider depends on the scale j and on the component f^m considered. We find in practice that $K = 1$ is sufficient to encode the differences between our components. Fixing $K = 1$, the final a priori model for each component f^m is : for a fixed scale j and a fixed orientation q, the neighborhoods

of wavelet coefficients $\mathbf{f}^m_{j,q,\overline{n}}$, for $n \in \mathbb{Z}^2$ are independent identically distributed with the same distribution as the Gaussian scale mixture $\sqrt{z^m_j}\ \mathbf{u}^m_{j,q}$, where the distribution of the multiplier z^m_j is independent of the orientation q.

Note that since the neighborhood of wavelet coefficients overlap for close locations, the independence can not hold in reality. However, our strategy is to retain from each estimated neighborhood $\mathbf{f}^m_{j,q,\overline{n}}$ only the central coefficient $f^m_{j,q,\overline{n}}$. Therefore, we do not need to make each estimated neighborhood consistent. Rather, we rely on the fact that neighborhoods themselves take into account statistical dependencies between coefficients, to ensure that the estimated coefficients $f^m_{j,q,\overline{n}}$ are consistent. The a priori model is then determined by the parameters of the Gaussian scale mixtures for each component f^m, each scale j and orientation q. We will describe how to choose these parameters in detail in section 3.3, but we first explain how the estimation will be carried out from this model.

3.2 Bayes least square estimate

In this section we explain how to compute the Bayes least square estimates of the neighborhood of coefficients for each component, given the a priori model we just described and the forward model for the observations $g^1,\ g^2,..,\ g^L$:

$$g^l = b^l\ *\ \Big[\ \sum_{m=1}^{M} a^{m,l}\ f^m\ \Big] + w^l \tag{3.4}$$

Here the beam functions b^l are known deterministic functions, the frequency dependencies $a^{m,l}$ are known scalars. The noise w^l is Gaussian and stationary, with known covariance, and is independent from one observation to the other.

To explain our estimation method, we break it down in several steps. We first explain the estimation of a single component by denoising a single observation. This follows closely [47]. Then we explain how to take the blurring into account for a single component. We derived this adaptation independently from the authors of the original paper who presented it succinctly in [48]. Here we give more details on the derivation of Bayes estimate for the problem of deblurring one observation ; in particular we explain the modeling assumptions made in this case. Then in subsection 3.2.3, we extend this method to the observations of several mixture of components, and show how to separate them.

3.2.1 Denoising one signal

Let us first consider the simple case where we observe one process polluted by noise : $g = f + w$. The equations for each single wavelet coefficient and for the neighborhood of wavelet coefficients read :

$$\begin{aligned} g_{j,q,\overline{n}} &= f_{j,q,\overline{n}} + w_{j,q,\overline{n}} \\ \mathbf{g}_{j,q,\overline{n}} &= \mathbf{f}_{j,q,\overline{n}} + \mathbf{w}_{j,q,\overline{n}} \end{aligned} \tag{3.5}$$

The Bayes least square estimate of the neighborhood $\mathbf{f}$ given the observed neighborhood $\mathbf{g}$ is the conditional expectation $E\{\mathbf{f}|\mathbf{g}\}$. The convenience of the representation of the neighborhood $\mathbf{f}$ by a Gaussian scale mixture is that given the multiplier z, Eq. (3.5) reduces to a sum of Gaussian vectors :

$$\mathbf{g} = \sqrt{z}\mathbf{u} + \mathbf{w} \tag{3.6}$$

When x and r are Gaussian vectors, the conditional expectation $E\{x|r\}$ of x given r, is :

$$E\{x|r\} = \mathbf{C}_{\mathbf{x},\mathbf{r}}(\mathbf{C}_{\mathbf{r}})^{-1}(r), \tag{3.7}$$

where $\mathbf{C}_{\mathbf{x},\mathbf{r}}$ is the covariance matrix between the vectors x and r. If in addition x and y are independent and $r = x + y$, then $\mathbf{C}_{\mathbf{x},\mathbf{r}} = \mathbf{C}_{\mathbf{x},\mathbf{x}+\mathbf{y}} = \mathbf{C}_{\mathbf{x},\mathbf{x}} + \mathbf{C}_{\mathbf{x},\mathbf{y}} = \mathbf{C}_{\mathbf{x},\mathbf{x}} \stackrel{def}{=} \mathbf{C}_{\mathbf{x}}$, and similarly $\mathbf{C}_{\mathbf{r}} = \mathbf{C}_{\mathbf{x}} + \mathbf{C}_{\mathbf{y}}$. The following result holds whenever x and y are two independent Gaussian vectors :

$$E\{x|x+y\} = \mathbf{C}_{\mathbf{x}}(\mathbf{C}_{\mathbf{x}} + \mathbf{C}_{\mathbf{y}})^{-1}(x+y) \tag{3.8}$$

Going back to Eq. (3.6), and using the independence of the Gaussian vectors $\mathbf{u}$ and $\mathbf{w}$, we obtain that conditioned on the random variable z :

$$E\{\mathbf{u}|\mathbf{g}, z\} = \sqrt{z}\,\mathbf{C}_{\mathbf{u}}(z\,\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}})^{-1}(\mathbf{g}), \tag{3.9}$$

using $\mathbf{C}_{\mathbf{u}} = \mathbf{C}_{\mathbf{f}}$, this leads to :

$$E\{\mathbf{f}|\mathbf{g}, z\} = z\,\mathbf{C}_{\mathbf{f}}(z\,\mathbf{C}_{\mathbf{f}} + \mathbf{C}_{\mathbf{w}})^{-1}(\mathbf{g}). \tag{3.10}$$

In other words the Bayes least square estimate of $\mathbf{f}$ given the observed vector $\mathbf{g}$ and given the multiplier z, is a Wiener filter applied to $\mathbf{g}$, the neighborhood of wavelet coefficients of the observation. Integrating the last equation with respect to the posterior distribution of the multiplier $p(z|\mathbf{g})$, we get the Bayes least square estimate of $\mathbf{f}$ given the observation $\mathbf{g}$:

$$E\{\mathbf{f}|\mathbf{g}\} = \int_0^\infty E\{\mathbf{f}|\mathbf{g}, z = z_0\}\; p(z = z_0|\mathbf{g})\; dz_0 \tag{3.11}$$

This estimate is a weighted average of the Wiener filters described in Eq. (3.10). The weights are determined by the posterior distribution, $p(z|\mathbf{g})$, which is computed via Bayes rule :

$$p(z = z_0|\mathbf{g}) = \frac{p(\mathbf{g}|z = z_0)p_z(z_0)}{\int p(\mathbf{g}|z = z')p_z(z')dz'}. \tag{3.12}$$

Here $p(\mathbf{g}|z = z')$ is a centered multidimensional Gaussian distribution with covariance matrix $z'\,\mathbf{C}_{\mathbf{f}} + \mathbf{C}_{\mathbf{w}}$, and p_z is the probability distribution of z (which we will describe in 3.3).

Following this procedure, one gets an estimate $E\{\mathbf{f}_{j,q,\overline{n}}|\mathbf{g}_{j,q,\overline{n}}\}$ for each neighborhood of coefficients $\mathbf{f}_{j,q,\overline{n}}$. One keeps only the central coefficient $f_{j,q,\overline{n}}$ of each of these estimated vector and reconstructs an estimate of the signal f by inverting the wavelet transform with these coefficients.

3.2.2 Deblurring one signal

We consider now the case where the observed signal is a blurred version of the one object : $g = f * b + w$. The convolution with the beam b correlates the signal spatially. As a result, the equations in wavelet space do not decouple any more :

$$g_{j,q,\overline{n}} = \left\langle f * b + w \ , \ \varphi^q_{j,\overline{n}} \right\rangle \tag{3.13}$$

$$g_{j,q,\overline{n}} = \left\langle f * b \ , \ \varphi^q_{j,\overline{n}} \right\rangle + w_{j,q,\overline{n}} \tag{3.14}$$

$$g_{j,q,\overline{n}} = \sum_{j',q',\overline{n}'} f_{j',q',\overline{n}'} \left\langle \varphi^{q'}_{j',\overline{n}'} * b \ , \ \varphi^q_{j,\overline{n}} \right\rangle + w_{j,q,\overline{n}} \tag{3.15}$$

Defining $b_{(j,q,\overline{n}),(j',q',\overline{n}')}$ by $b_{(j,q,\overline{n}),(j',q',\overline{n}')} = \left\langle \varphi^{q'}_{j',\overline{n}'} * b \ , \ \varphi^q_{j,\overline{n}} \right\rangle$, we obtain :

$$g_{j,q,\overline{n}} = \sum_{j',q',\overline{n}'} b_{(j,q,\overline{n}),(j',q',\overline{n}')} f_{j',q',\overline{n}'} + w_{j,q,\overline{n}} \tag{3.16}$$

Therefore, a particular neighborhood $\mathbf{f}_{j',q',\overline{n}',K}$ contributes to every observed wavelet coefficient $g_{j,q,\overline{n}}$. In theory, one would obtain the best estimate of $\mathbf{f}_{j',q',\overline{n}',K}$ by using the information in every wavelet coefficients $g_{j,q,\overline{n}}$. This would be a very difficult estimation problem, moreover, our final goal is not the estimation of the neighborhoods themselves but rather their central coefficients. So we do not intend to use the full set of coefficients $g_{j,q,\overline{n}}$ to estimate each neighborhood. Rather, by considering the properties of the beam and the wavelets, we claim that using only the observed neighborhood $\mathbf{g}_{j,q,\overline{n},K}$ yields a sufficiently good estimation of the object neighborhood $\mathbf{f}_{j,q,\overline{n},K}$, when K is chosen appropriately.

To see that, let us fix an index $j, q, \overline{n}$ and consider the coefficients $b_{(j',q',\overline{n}'),(j,q,\overline{n})}$ for all $j', q', \overline{n}'$. Using the fact that the beam is radially symmetric, we can rewrite these coefficients :

$$b_{(j,q,\overline{n}),(j',q',\overline{n}')} = \left\langle \varphi^{q'}_{j',\overline{n}'} * b \ , \ \varphi^q_{j,\overline{n}} \right\rangle \tag{3.17}$$

$$b_{(j,q,\overline{n}),(j',q',\overline{n}')} = \left\langle \varphi^{q'}_{j',\overline{n}'} \ , \ b * \varphi^q_{j,\overline{n}} \right\rangle \tag{3.18}$$

$$b_{(j,q,\overline{n}),(j',q',\overline{n}')} = \left\langle \widehat{\varphi^{q'}_{j',\overline{n}'}} \ , \ \widehat{b * \varphi^q_{j,\overline{n}}} \right\rangle \tag{3.19}$$

$$b_{(j,q,\overline{n}),(j',q',\overline{n}')} = \int \widehat{b}(\xi) \ \widehat{\varphi^q_{j,\overline{n}}}(\xi) \ \overline{\widehat{\varphi^{q'}_{j',\overline{n}'}}}(\xi) \ d\xi \tag{3.20}$$

Most of these coefficients are really small :

1. If $|j - j'|$ is large, then, since the wavelet is well localized in frequency, $\varphi^q_{j,\overline{n}}$ and $\varphi^{q'}_{j',\overline{n}'}$ are concentrated in different frequency bands. Hence $\int |\widehat{\varphi^q_{j,\overline{n}}}(\xi)||\overline{\widehat{\varphi^{q'}_{j',\overline{n}'}}}(\xi)| d\xi$ is small and by Eq. (3.20), $b_{(j,q,\overline{n}),(j',q',\overline{n}')}$ is small.
2. If $|q - q'|$ is large, then, since oriented wavelet are localized in different parts of the frequency plane, again the support of $\widehat{\varphi^q_{j,\overline{n}}}$ and $\widehat{\varphi^{q'}_{j',\overline{n}'}}$ are different. Hence $\int |\widehat{\varphi^q_{j,\overline{n}}}(\xi)| \ |\overline{\widehat{\varphi^{q'}_{j',\overline{n}'}}}(\xi)| \ d\xi$ is small and by Eq. (3.20), $b_{(j,q,\overline{n}),(j',q',\overline{n}')}$ is small.

3. If $|\overline{n} - \overline{n}'|$ is large, we use the localization in space of both the beam b and the wavelet to argue that $b_{(j,q,\overline{n}),(j',q',\overline{n}')}$ is small. We define the width of a function h by the minimal length of the interval I such that : $\int_I |h(x)^2|dx > \eta \int |h(x)^2|dx$, fixing $\eta = .9$ for example. If $|2^j\overline{n} - 2^{j'}\overline{n}'| > |b| + (2^{j-1} + 2^{j'-1})l$ where l is the width of the wavelet and $|b|$ the width of the beam, the support of the functions $\varphi^{q'}_{j',\overline{n}'}$ and $b * \varphi^{q}_{j,\overline{n}}$ essentially do not intersect, so that by Eq. (3.18), $b_{(j,q,\overline{n}),(j',q',\overline{n}')}$ is small.

Note that since the wavelets we use here are compactly supported in frequency (cf. section 4.3 in Chapter 4 for the details), we actually have : $b_{(j,q,\overline{n}),(j',q',\overline{n}')} = 0$ for $|j - j'| > 1$ or $|q - q'| > 1$. Hence we argue that $b_{(j,q,\overline{n}),(j',q',\overline{n}')}$ is not significant unless $|j - j'| \leq 1$, $|q - q'| \leq 1$ and $|2^j\overline{n} - 2^{j'}\overline{n}'| > |b| + (2^{j-1} + 2^{j'-1})l$. It turns out that practically, the cross terms $b_{(j,q,\overline{n}),(j',q',\overline{n}')}$ for different orientations $q' = q+1$ or $q' = q-1$ are negligible as well. As far as the scale $j' = j+1$ or $j' = j-1$ is concerned, the coefficients $b_{(j,q,\overline{n}),(j',q',\overline{n}')}$ are in practice smaller than the coefficients at the same scale $b_{(j,q,\overline{n}),(j,q,\overline{n}')}$ unless $n = n'$.

Putting this together, we obtain that the contribution of a particular wavelet coefficient $f_{j,q,\overline{n}}$ is most important in the neighborhood of observed coefficients of the form $\mathbf{g}_{j,q,\overline{n},K_b}$ where $K_b = 2^{-j}|b| + l$. Keeping in mind that we will retain only the central coefficient $f_{j,q,\overline{n}}$ from the estimated neighborhood $\mathbf{f}_{j,q,\overline{n},K_f}$, (where K_f is the size of the neighborhood needed to capture the spatial dependences of the wavelet coefficients of f), it is then reasonable to use only the observed neighborhood $\mathbf{g}_{j,q,\overline{n},K}$ to estimate $\mathbf{f}_{j,q,\overline{n},K_f}$ choosing $K = \max\{K_f, K_b\}$.

Using Eq. (3.16) for each coefficient in the neighborhood $\mathbf{g}_{j,q,\overline{n},K}$, we get :

$$\mathbf{g}_{j,q,\overline{n},K} = B_{j,q,\overline{n},K}\mathbf{f}_{j,q,\overline{n},K} + R_{j,q,\overline{n},K} + \mathbf{w}_{j,q,\overline{n},K} \tag{3.21}$$

with :

$$B_{j,q,\overline{n},K} = \left\{b_{(j_1,q_1,\overline{n}_1),(j_2,q_2,\overline{n}_2)}\right\}_{(j_1,q_1,\overline{n}_1),(j_2,q_2,\overline{n}_2)\in\mathcal{V}_{j,q,\overline{n},K}} \tag{3.22}$$

$$R_{j,q,\overline{n},K} = \sum_{\substack{(j_1,q_1,\overline{n}_1)\notin\mathcal{V}_{j,q,\overline{n},K} \\ \text{or } (j_2,q_2,\overline{n}_2)\notin\mathcal{V}_{j,q,\overline{n},K}}} b_{(j_1,q_1,\overline{n}_1),(j_2,q_2,\overline{n}_2)}\ f_{j_2,q_2,\overline{n}_2} \tag{3.23}$$

Here, we have separated the different contributions to the observed neighborhood $\mathbf{g}_{j,q,\overline{n},K}$ into three terms : the contribution of the same neighborhood in the object $B_{j,q,\overline{n},K}\mathbf{f}_{j',q',\overline{n}',K}$, the contribution of the same neighborhood in the noise $\mathbf{w}_{j,q,\overline{n},K}$ and the contribution from remaining wavelet coefficients in the object $R_{j,q,\overline{n},K}$.

As we saw earlier, the coefficients $b_{(j_1,q_1,\overline{n}_1),(j_2,q_2,\overline{n}_2)}$ that appear in $R_{j,q,\overline{n},K}$ are rather small and therefore the contribution of this term can be considered negligible. We will consider this term as additional noise and work with the model :

$$\mathbf{g}_{j,q,\overline{n},K} = B_{j,q,\overline{n},K}\ \mathbf{f}_{j,q,\overline{n},K} + \mathbf{w}'_{j,q,\overline{n},K} \tag{3.24}$$

where $B_{j,q,\overline{n},K}$ is the matrix described in Eq. (3.22), w' is modeled by Gaussian noise and $\mathbf{f}_{j,q,\overline{n},K}$ by a Gaussian scale mixture.

Dropping the indexes and using Eq. (3.7), we find that the expected value of the neighborhood $\mathbf{f}$ given the observed neighborhood $\mathbf{g}$ and the multiplier z is the Wiener filter :

$$E\{\mathbf{f}|\mathbf{g},z\} = z\,\mathbf{C_f}\,B^*(zB\,\mathbf{C_f}\,B^* + \mathbf{C_w}')^{-1}\mathbf{g} \tag{3.25}$$

The full Bayes least square is again a weighted sum of these filters, with the weights given by the posterior distribution $p(z|\mathbf{g})$ computed via Eq. (3.12). The prior $p(\mathbf{g}|z)$ also takes the blurring into account : it is a multidimensional Gaussian centered and with covariance matrix $zB\,\mathbf{C_f}\,B^* + \mathbf{C_w}'$. As before, only the central coefficient of each estimated neighborhood $\mathbf{f}$ is used to reconstruct the object f via the inverse wavelet transform.

With this procedure in mind, we can now explain how to extend this method to the problem of separation of blurred mixtures of signals.

3.2.3 Separating blurred mixtures of signals

Given the model for the mixture of components in Eq. (3.4), the analog to equation (3.16) is :

$$\forall l \in [\![1,L]\!], \quad g^l_{j,q,\overline{n}} = \sum_{m=1}^{M} \sum_{j',q',\overline{n}'} a^{m,l}\; b^l_{(j,q,\overline{n}),(j',q',\overline{n}')} f^m_{j',q',\overline{n}'} + w^l_{j,q,\overline{n}} \tag{3.26}$$

As we argued before, most of the coefficients $b^l_{(j,q,\overline{n}),(j',q',\overline{n}')}$ are very small. Therefore, the influence of a particular wavelet coefficient of object m_o, $f^{m_o}_{j,q,\overline{n}}$, will be mostly seen in the neighborhood $\mathbf{g}^l_{j,q,\overline{n},K^l}$ of each observation g^l. However, this time, the contribution of object m_o is not the only significant contribution in $\mathbf{g}^l_{j,q,\overline{n},K^l}$: each component f^m potentially gives such a significant contribution. It is then natural consider the L neighborhoods $\mathbf{g}^l_{j,q,\overline{n},K^l}$, for $l = 1,..L$ in conjunction to estimate at the same time the M neighborhoods $\mathbf{f}^m_{j',q',\overline{n}',K^l}$, for $m = 1,..M$. Note that the size K^l of the observed neighborhoods $\mathbf{g}^l_{j,q,\overline{n},K^l}$ we have to consider depends on the beam size for observation l, whereas the size of the neighborhood $\mathbf{f}^m_{j',q',\overline{n}',K^m}$ that is needed to describe the spatial coherence of the wavelet coefficients of object m, depends on the object itself. As before, we will choose K to be the maximum of these parameters : $K = \max_{l\in[\![1,L]\!],\, m\in[\![1,m]\!]}\{K^l, K^m\}$. This way, all the neighborhoods we consider for a fixed scale j have the same size.

Separating again significant from non-significant contributions, we get :

$$\forall l \in [\![1,L]\!], \quad \mathbf{g}^l_{j,q,\overline{n},K} = B^l_{j,q,\overline{n},K}\Big(\sum_{m=1}^{M} a^{m,l}\; \mathbf{f}^m_{j,q,\overline{n},K}\Big) + \mathbf{w}^{,\,l}_{j,q,\overline{n},K} \tag{3.27}$$

with :

$$B^l_{j,q,\overline{n},K} = \big\{b^l_{(j_1,q_1,\overline{n}_1),(j_2,q_2,\overline{n}_2)}\big\}_{(j_1,q_1,\overline{n}_1),(j_2,q_2,\overline{n}_2)\in\mathcal{V}^2_{j,q,\overline{n},K}} \tag{3.28}$$

$$\mathbf{w}^{,\,l}_{j,q,\overline{n},K} = \mathbf{w}^l_{j,q,\overline{n},K} + \sum_{\substack{(j_1,q_1,\overline{n}_1)\notin\mathcal{V}_{j,q,\overline{n},K} \\ \text{or } (j_2,q_2,\overline{n}_2)\notin\mathcal{V}_{j,q,\overline{n},K}}} b^l_{(j_1,q_1,\overline{n}_1),(j_2,q_2,\overline{n}_2)}\Big(\sum_{m=1}^{M} a^{m,l}\; f^m_{j_2,q_2,\overline{n}_2}\Big) \tag{3.29}$$

Let us fix the neighborhood $\mathcal{V}_{j,q,\overline{n},K}$ that we consider. The matrices $B^l_{j,q,\overline{n},K}$ for $l \in [\![1, L]\!]$ and the frequency dependences $a^{m,l}$ for $l \in [\![1, L]\!]$ and $m \in [\![1, M]\!]$ are deterministic and known. Each vector $\mathbf{w}^{,l}_{j,q,\overline{n},K}$, for $l \in [\![1, L]\!]$ is supposed to be Gaussian, centered, with known covariance matrix. Each vector $\mathbf{f}^m_{j,q,\overline{n},K}$ follows the distribution of a Gaussian scale mixture $\sqrt{z^m}\ \mathbf{u}^m$ for each m in $[\![1, M]\!]$. The noise terms are independent from one observation to another. Moreover, the objects are also assumed to be independent from each other and from the noise.

To derive the Bayes least square estimate under this model, it is useful to consider the observed neighborhoods as constituents of a larger vector G :

$$\begin{aligned} G &= (\mathbf{g}^1_{j,q,\overline{n},K}, \mathbf{g}^2_{j,q,\overline{n},K}, \dots, \mathbf{g}^L_{j,q,\overline{n},K}) && (3.30)\\ G &= (g^1_{i_1}, g^1_{i_2}, \dots, g^2_{i_1}, g^2_{i_2}, \dots\dots, g^L_{i_1}, g^L_{i_2}, \dots), \text{ with } i_j \in \mathcal{V}_{j,q,\overline{n},K} && (3.31)\end{aligned}$$

Similarly, we stack the noise neighborhoods into a larger vector W :

$$W = (w^1_{i_1}, w^1_{i_2}, \dots, w^2_{i_1}, w^2_{i_2}, \dots\dots, w^L_{i_1}, w^L_{i_2}, \dots), \text{ with } i_j \in \mathcal{V}_{j,q,\overline{n},K} \quad (3.32)$$

And the objects neighborhoods into a larger vector F :

$$F = (f^1_{i_1}, f^1_{i_2}, \dots, f^2_{i_1}, f^2_{i_2}, \dots\dots, f^M_{i_1}, f^M_{i_2}, \dots), \text{ with } i_j \in \mathcal{V}_{j,q,\overline{n},K} \quad (3.33)$$

This way, the L equations in Eq. (3.27) can be written as a single matrix equation :

$$G^T = DF^T + W^T \quad (3.34)$$

where D is the matrix :

$$D = \begin{pmatrix} a^{1,1}\ B^1_{j,q,\overline{n},K} & a^{1,2}\ B^1_{j,q,\overline{n},K} & \cdots & a^{1,M}\ B^1_{j,q,\overline{n},K} \\ a^{2,1}\ B^2_{j,q,\overline{n},K} & a^{2,2}\ B^2_{j,q,\overline{n},K} & \cdots & a^{2,M}\ B^2_{j,q,\overline{n},K} \\ \vdots & \vdots & \ddots & \vdots \\ a^{L,1}\ B^L_{j,q,\overline{n},K} & a^{L,2}\ B^L_{j,q,\overline{n},K} & \cdots & a^{L,M}\ B^L_{j,q,\overline{n},K} \end{pmatrix}, \quad (3.35)$$

where each $a^{m,l}\ B^l_{j,q,\overline{n},K}$ is a block of size $L \times L$, with $L = |\mathcal{V}_{j,q,\overline{n},K}|$ the cardinal of the neighborhood $\mathcal{V}_{j,q,\overline{n},K}$. Writing the equation in matrix form makes the computation of the estimator very similar to what we saw in section 3.2.2, with the exception the the "object" vector F is not a simple scale mixture of Gaussians, but takes in account M multipliers :

$$\begin{aligned} F &\overset{dist.}{\equiv} (\sqrt{z^1}\ u^1_{i_1}, \sqrt{z^1}\ u^1_{i_2}, \dots, \sqrt{z^2}\ u^2_{i_1}, \sqrt{z^2}\ u^2_{i_2}, \dots\dots \sqrt{z^M}\ u^M_{i_1}, \sqrt{z^M}\ u^M_{i_2}, \dots) \\ F &\overset{dist.}{\equiv} \sqrt{Z}\ \circ\ U && (3.36)\end{aligned}$$

where U is a Gaussian vector, Z contains each multiplier z^m repeated $|\{\mathcal{V}_{j,q,\overline{n},K}\}|$ times and $\circ$ denotes the multiplication coordinate by coordinate.

Using Eq. (3.7), we obtain formally the conditional expectation of F given the observation G and the M multipliers $\{z^m\}_m$:

$$E\{F|G, z^1, z^2, \dots, z^M\} = \mathbf{C}_{\sqrt{Z}\,\circ\,U}\, D^* \Big(D\, \mathbf{C}_{\sqrt{Z}\,\circ\,U}\, D^* + \mathbf{C}_{\mathbf{W}} \Big)^{-1} G \quad (3.37)$$

The Bayes least square estimate of the F given the observations is then :

$$E\{F|G\} = \int_{\mathbb{R}_+^M} E\{F|G, z^1, z^2, \ldots, z^M\}\; p(z^1, z^2, \ldots, z^M|G)\; dz^1 dz^2 \ldots dz^M \quad (3.38)$$

The posterior is as usual obtained via Bayes rule :

$$p(z^1, z^2, \ldots, z^M|G) = \frac{p(G|z^1, z^2, \ldots, z^M) \prod_{m=1}^M p_{z^m}(z^m)}{\int_{\mathbb{R}_+^M} p(G|z^1 = \alpha^1, z^2 = \alpha^2, \ldots, z^M = \alpha^M) \prod_{m=1}^M p_{z^m}(\alpha^m)} \quad (3.39)$$

with p_{z^m} is the distribution of the multiplier corresponding to the object f^m and the prior distribution for the observed vector G, $p(G|z^1, z^2, \ldots, z^M)$, is again a multidimensional Gaussian, centered and with covariance matrix $\mathcal{C}_Z = \Big(D\, \mathbf{C}_{\sqrt{Z} \circ U}\, D^* + \mathbf{C_W}\Big)$.

We shall now relate these equations involving the abstract vectors G, F and W to our original neighborhoods of wavelet coefficients. Since the noise terms w^l for each observation are independent from each other, the covariance matrix $\mathbf{C_W}$ is block diagonal, with L blocks. Each block is the covariance matrix of the l^{th} noise neighborhood $\mathbf{w}^l_{j,q,\overline{n},K}$: $\mathbf{C}_{\mathbf{w^l_{j,q,\overline{n},K}}}$. The covariance matrix $\mathbf{C}_{\sqrt{\mathbf{Z}} \circ \mathbf{U}}$ is also block diagonal because the objects f^m, $m = 1, .., M$ are independent from each other. It is constituted by M blocks, each of which is the covariance matrix of an object neighborhood $\mathbf{f}^m_{j,q,\overline{n},K}$ times the multiplier z^m, i.e. $z^m\, \mathbf{C}_{\mathbf{f^m_{j,q,\overline{n},K}}}$. (The value z^m appears here in the covariance matrix because $\mathbf{C}_{\sqrt{\mathbf{Z}} \circ \mathbf{U}}$ was computed conditionally on the multipliers.) The covariance matrix $\mathcal{C}_Z$ is defined by blocks $\mathcal{C}_Z = \{\mathcal{C}_Z(l_1, l_2)\}_{\{l_1, l_2 \in [\![1, L]\!]^2\}}$ with :

$$\mathcal{C}_Z(l_1, l_2) = \sum_{m=1}^M z^m a^{m,l_1} \overline{a^{m,l_2}}\; B^{l_1}_{j,q,\overline{n},K}\, \mathbf{C}_{\mathbf{f^m_{j,q,\overline{n},K}}}\, B^{l_2\ *}_{j,q,\overline{n},K} + \delta_{\{l_1 = l_2\}}\, \mathbf{C}_{\mathbf{w^{l_1}_{j,q,\overline{n},K}}} \quad (3.40)$$

As a result, the prior $p(G|z^1, z^2, \ldots, z^M)$ reads :

$$p(G|z^1, z^2, \ldots, z^M) = \tfrac{1}{(2\pi)^{LV/2} \det(\mathcal{C}_Z)}\ \exp\Big\{-\frac{G{\mathcal{C}_Z}^{-1} G^T}{2}\Big\}, \quad (3.41)$$

where V is the cardinal of the neighborhood considered. The conditional expectation of the neighborhood $\mathbf{f}^m_{j,q,\overline{n},K}$ given the multipliers and the observed neighborhoods is :

$$E\{\mathbf{f}^m_{j,q,\overline{n},K}|G, z^1, z^2, \ldots, z^M\} = \sum_{l=1}^L z^m\ a^{m,l}\ \mathbf{C}_{\mathbf{f^m_{j,q,\overline{n},K}}}\, B^{l\ *}_{j,q,\overline{n},K} \Big({\mathcal{C}_Z}^{-1} G^T\Big)^l \quad (3.42)$$

This is integrated with respect to the posterior distribution of the M-uple of multipliers $(z^1, z^2, .., z^M)$ to find the Bayes least square estimate of $\mathbf{f}^m_{j,q,\overline{n},K}$ given the observed neighborhoods $\mathbf{g}^l_{j,q,\overline{n},K}$ grouped in the vector G :

$$E\{\mathbf{f}^m_{j,q,\overline{n},K}|G\} = \tfrac{1}{C(G)} \int_{\mathbb{R}_+^M} E\{\mathbf{f}^m_{j,q,\overline{n},K}|G, z^1, z^2, \ldots, z^M\} e^{-\frac{G{\mathcal{C}_Z}^{-1} G^T}{2}} \prod_{m=1}^M \big[p_{z^m}(z^m)\; dz^m\big] \quad (3.43)$$

with

$$C(G) = \int_{\mathbb{R}_+^M} e^{-\frac{G\mathcal{C}_Z{}^{-1}G^T}{2}} \prod_{m=1}^{M} \left[p_{z^m}(z^m)\ dz^m\right] \tag{3.44}$$

Note that Eq.(3.43) and (3.44) are M-fold integrations.

In the next section, we describe how we obtain the parameters necessary to compute these estimations. These are the covariance matrices of the objects and noises neighborhoods, as well as the probability densities for the multipliers.

3.3 Choice of the parameters

As explained in 3.1.3, for a fixed value of K, we assume that for each scale j, orientation q, and component f^m, the neighborhoods of wavelets coefficients $\{f^m_{j,q,\overline{n}}\}_{\overline{n}\in Z^2}$ follow a scale mixture of Gaussian $\sqrt{z^m_j}\ \mathbf{u}^m_{j,q}$, where the distribution of the multiplier is independent of the orientation. Moreover, we assumed that each noise maps w^l is modeled by a stationary process. Therefore neither the covariance matrices nor the multipliers actually depend on the location $\overline{n}$. To compute the Bayes estimation described above, we need : the noise covariances $\mathbf{C}_{\mathbf{w^l_{j,q,0,K}}}$, the component covariances $\mathbf{C}_{\mathbf{f^m_{j,q,0,K}}}$, and the probability distributions $p_{z^m_j}$ for $l \in [\![1, L]\!]$, $m \in [\![1, M]\!]$, for all scales j and all orientations q. (As we saw in the previous section, the size of the neighborhood K is the same for all observations and components.)

3.3.1 Covariance matrices of the noise neighborhoods

We assume in this work that the noise term w^l for each observation g^l, $l \in [\![1, L]\!]$, is Gaussian and stationary. It can be white, and in this case, we assume that we have an estimate of the standard deviation σ^l for each l. The noise could also be colored, and in that case, we assume that we know its spatial covariance matrix noted $\mathbf{C^s_{w^l}}$ (where $\mathbf{C^s_{w^l}}(\overline{x} - \overline{x}') = \mathbf{Cov}\big(w^l(\overline{x}), w^l(\overline{x}')\big)$, for any $\overline{x}$ in $\mathbb{R}^2$ and $\overline{x}'$ in $\mathbb{R}^2$). When the noise is white, $\mathbf{C^s_{w^l}}(\overline{x}) = (\sigma^l)^2\delta_{\overline{x}=0}$. The covariance matrices of the neighborhoods of the noise terms w^l are by definition :

$$\mathbf{C}_{\mathbf{w^l_{j,q,0,K}}} = \left\{\mathbf{Cov}(w^l_{j_1,q_1,\overline{n}_1}, w^l_{j_2,q_2,\overline{n}_2})\right\}_{\{(j_1,q_1,\overline{n}_1),(j_2,q_2,\overline{n}_2)\in\mathcal{V}^2_{j,q,0,K}\}} \tag{3.45}$$

Suppose we use two-dimensional wavelet transform with Q orientations. We note T the wavelet transform operator :

$$T : \begin{array}{ccc} L^2(\mathbb{R}^2) & \rightarrow & l^2(\mathbb{Z}^3 \times [\![1, Q]\!]) \\ h & \mapsto & \left\{\left\langle h\ ,\ \varphi^q_{j,\overline{n}}\right\rangle\right\}_{j\in\mathbb{Z},\overline{n}\in\mathbb{Z}^2,q\in[\![1,Q]\!]} \end{array} \tag{3.46}$$

Then $w^l_{j,q,\overline{n}} = \left\langle w^l\ ,\ \varphi^q_{j,\overline{n}}\right\rangle = \{T(w^l)\}_{j,q,\overline{n}}$. Since T is linear, then :

$$E\{w^l_{j_1,q_1,\overline{n}_1}\} = \left\{T\big(E\{w^l\}\big)\right\}_{j,q,\overline{n}} \tag{3.47}$$

$$\mathbf{Cov}(w^l_{j_1,q_1,\overline{n}_1}, w^l_{j_2,q_2,\overline{n}_2}) = \left\{T\,\mathbf{C^s_{w^l}}\,T^*\right\}_{(j_1,q_1,\overline{n}_1),(j_2,q_2,\overline{n}_2)} \tag{3.48}$$

w^l is centered so $E\{w^l\} = 0$ and therefore $E\{w^l_{j_1,q_1,\overline{n}_1}\} = 0$. The covariance terms can be written in term of scalar products of one wavelet with another, modulated by the covariance $\mathbf{C}^{\mathbf{s}}_{\mathbf{w^l}}$:

$$\begin{aligned}\mathbf{Cov}(w^l_{j_1,q_1,\overline{n}_1}, w^l_{j_2,q_2,\overline{n}_2}) &= \left\langle \varphi^{q_1}_{j_1,\overline{n}_1}, \mathbf{C}^{\mathbf{s}}_{\mathbf{w^l}}\, \varphi^{q_2}_{j_2,\overline{n}_2} \right\rangle_{L^2(\mathbb{R}^2)} \\ \mathbf{Cov}(w^l_{j_1,q_1,\overline{n}_1}, w^l_{j_2,q_2,\overline{n}_2}) &= \int_{\mathbb{R}^2\times\mathbb{R}^2} \overline{\varphi^{q_1}_{j_1,\overline{n}_1}}(\overline{x})\, \mathbf{C}^{\mathbf{s}}_{\mathbf{w^l}}(\overline{x}-\overline{x}')\varphi^{q_2}_{j_2,\overline{n}_2}(\overline{x}')\; d\overline{x}d\overline{x}'\end{aligned} \tag{3.49}$$

When the noise is white, this reduces to :

$$\mathbf{Cov}(w^l_{j_1,q_1,\overline{n}_1}, w^l_{j_2,q_2,\overline{n}_2}) = (\sigma^l)^2 \left\langle \varphi^{q_1}_{j_1,\overline{n}_1}, \varphi^{q_2}_{j_2,\overline{n}_2} \right\rangle_{L^2(\mathbb{R}^2)} \tag{3.50}$$

Hence, the covariance matrices of the noise neighborhoods can be computed prior to computing the estimates, if the wavelet transform, the size of the neighborhoods and the spatial covariances of the noises are known beforehand.

3.3.2 Covariance matrices of the objects neighborhoods

In the case of the deblurring of a single object, Portilla et al. propose in [47] a method to estimate the covariance of the single object from the observation itself. This method is based on the fact that the covariance of a signal h is the inverse Fourier transform of its spectral power $P_h = |\hat{h}|^2$, and that the spectral power of two independent signals is the sum of their spectral powers. Computing the spectral powers in the case of one blurred component : $g = f * b + w$, one gets $P_g = P_{b*f} + P_w$. The spectral power of the convolution $b * f$ is $P_{b*f} = |\hat{b}|^2 P_f$. One can then estimate P_f knowing P_g from the observation and P_w for the noise, being careful to regularize the division by $|\hat{b}|^2$, as is explained in [47].

We extend this procedure to the case of blurred mixtures of components defined by Eq. (3.4). The power spectral densities now are :

$$\forall l \in [\![1, L]\!], \quad P_{g^l} = |\hat{b^l}|^2 \big(\sum_{m=1}^{M} |a^{m,l}|^2 P_{f^m}\big) + P_{w^l}. \tag{3.51}$$

Using the method proposed in [47], we can estimate the L linear combinations $S^l = \sum_{m=1}^{M} |a^{m,l}|^2 P_{f^m}$. If the matrix $A = \{|a^{m,l}|^2\}_{m\in[\![1,M]\!], l\in[\![1,L]\!]}$ is well conditioned, then we can recover the P_{f^m} using the pseudo-inverse A^*A and keeping only the positive part :

$$\forall m \in [\![1, M]\!], \quad P_{f^m}(\xi) = \Big[\Big\{(A^*A)^{-1}A^*\big(S^1(\xi), S^2(\xi), .., S^L(\xi)\big)^T\Big\}^m\Big]_+ \tag{3.52}$$

It turns out that this method is not well suited to our astrophysical problem for several reasons. The frequency dependence of the Galaxy dust (component f^4) and the point sources (component f^3) are very close in the range of frequency of our observed data. (Typically $|a(3,l) - a(4,l)| < 10^{-2}|a(4,l)|$.) Hence we are not able to separate their power spectrum with this method. Moreover, we made up test cases where we

considered only the CMB component and the clusters of galaxies component. In these cases, the method should technically work. (A is then well conditioned). However in practice, we find that the power spectrum of the clusters of galaxies is negligible compared to that of the noise and of the CMB. Therefore, we were not able to estimate it precisely enough with this method. (In fact, it would most of the time be estimated to 0 by taking the positive part in Eq. (3.52).)

One could imagine that another method of estimation, using only the observations but in a different way, may be able to solve the problem for the clusters of galaxies. However, this is not the case for the first problem we pointed out. When the frequency dependences of two objects are equal, they are formally merged into a single component from the point of view of Eq. (3.4). Therefore, one can not distinguish these components, or any of their features, based solely on the observations g^l and Eq. (3.4). A priori knowledge on the components has to be used in addition to the Eq. (3.4), even for the estimation of the covariance matrices. To our knowledge, there is no physical quantity well understood by astrophysicists for each of the components we consider and that we can use to constrain the estimation of the covariance matrices. Since we have at hand numerical simulations of each of the components we consider, we use them to compute templates for the covariance matrices of the neighborhoods of wavelet coefficients $\mathbf{C}_{\mathbf{f^m_{j,q,0}}}$.

Note that in practice, the neighborhood covariance matrices (for the components and also the noise terms), depend on the wavelet used and on the resolution of the observed data. The dependence on the wavelet is clear since each term in the covariance matrix of a neighborhood involves the wavelet itself (as we saw in Eq. (3.49)). The resolution of the observed data, i.e. the physical size of a pixel in the observed image, determines the physical size of the finest scale of the wavelet transform applied to this image. Therefore, when considering different experimental conditions, there is no reason why the abstract wavelet scales j of the computed wavelet transform always correspond to the same or similar physical scales. As a consequence, for each experiment, we will have to recompute the template covariance matrices of the neighborhoods for each component and for each noise term.

3.3.3 Prior distribution of the multipliers

We shall now describe how to determine the prior distributions of the multipliers z_j^m. The Gaussian scale mixture model imposes only two restrictions on the choice of the probability distribution $p_{z_j^m}$ which are : $p_{z_j^m}$ should be supported in $\mathbb{R}^+$ (that is $z_j^m \geq 0$), and its first moment should be 1 (i.e. $E\{z_j^m\} = 1$). Any choice of $p_{z_j^m}$ that satisfies these conditions is technically valid, so we have to consider the properties of the component f^m to make a choice.

When the component f^m is well modeled by a Gaussian process, the distribution of its wavelet coefficients at each scale is also Gaussian. Hence the neighborhoods $\mathbf{f}^m_{j,q,\overline{n},K}$ are well modeled by Gaussian vectors, in which case the multipliers z_j^m should not be used. As a result, if the component f^m is known to be well modeled by a Gaussian process, the distribution of the multipliers should be set to $p_{z_j^m}(x) = \delta_{\{x=1\}}$

for each scale j.

In the other cases, i.e. when f^m is not well modeled by a Gaussian process, or when this information is not available a priori, the choice has to be made on the basis of the empirical distributions of the wavelet neighborhoods $\mathbf{f}^m_{j,q,\overline{n},K}$. In order to obtain the most accurate model, one would ideally want to solve for the distribution $p_{z_j^m}$ using the empirical joint distribution of the neighborhoods vectors $\mathbf{f}^m_{j,q,\overline{n},K}$. A maximum likelihood approach to estimate $p_{z_j^m}$ for the problem of denoising a natural image has been proposed in [60]. However the authors argue that this estimation does not yield better estimates than Jeffrey's non-informative prior which in this case is a uniform probability on the logarithm of z :

$$p_{\log z}(u) = \frac{1}{V_{max} - V_{min}} \; \delta_{\{V_{min} \leq u \leq V_{max}\}} \tag{3.53}$$

i.e :

$$p_z(u) = \frac{1}{V_{max} - V_{min}} \; \delta_{\{V_{min} \leq \log u \leq V_{max}\}} \; \frac{1}{u} \tag{3.54}$$

where V_{min} and V_{max} are chosen so that $-\infty < V_{min} < V_{max} < \infty$.

Computing these estimations from the full neighborhoods for each component f^m would be computationally very costly in our case. Moreover, we find that using Jeffrey's prior when the neighborhood can not be considered Gaussian leads to good first estimates of our components. When we want to refine the model to obtain a better estimate for the component f^m, we choose to fit a prior $p_{z_j^m}$ considering only the marginal distribution of the central coefficient in the neighborhood $\mathbf{f}^m_{j,q,\overline{n},K}$. This amounts to deriving numerically the distribution $p_{z_j^m}$, considering the empirical distribution of the set of all the wavelet coefficients $\{f^m_{j,q,\overline{n}}\}_{\overline{n} \in \mathbb{Z}^2, q \in [\![1,Q]\!]}$ of the template component f^m at scale j, and the one-dimensional Gaussian scale mixture model :

$$\forall \overline{n} \in \mathbb{Z}^2, \forall \; q \in [\![1, Q]\!], \quad f^m_{j,q,\overline{n}} \overset{dist.}{\equiv} \sqrt{z_j^m} \; u, \tag{3.55}$$

where u is a scalar Gaussian random variable, centered and of variance $(\sigma_j^m)^2$. (Note that this variance was computed in the previous section as part of the covariance matrix $\mathbf{C}_{\mathbf{f_{j,q,0,K}}}$). Let us explain our

ad-hoc procedure for the derivation of the prior

with a formal one-dimensional Gaussian scale mixture : $x = \sqrt{z} \; u$, where all the random variables are scalar, and u is Gaussian (centered, variance σ^2). Taking the logarithm of the absolute values yields

$$\log |x| = \frac{1}{2} \log z + \log |u|, \tag{3.56}$$

from which we derive the relation between the probability densities $p_{\log|x|}$, $p_{\log z}$ and $p_{\log|u|}$:

$$p_{\log|x|}(v) = \big(\, p_{\frac{1}{2}\log z}(\cdot) * p_{\log|u|}(\cdot) \, \big)(v) \tag{3.57}$$

$$p_{\log|x|}(v) = \big(\, 2 \, p_{\log z}(2\,\cdot) * p_{\log|u|}(\cdot) \, \big)(v) \tag{3.58}$$

where $*$ denotes convolution and

$$p_{\log|u|}(v) = \frac{2}{\sqrt{2\pi\sigma^2}} e^{-\frac{e^{2v}}{2\sigma^2}+v} \tag{3.59}$$

At this point the author of [60], propose to deconvolve Eq. (3.58) and fit a Gaussian to the result, thus assuring that the estimated prior is a proper probability distribution. As a result, they restrict themselves to a log-normal distribution for the multiplier z. We take a different approach, not fitting a Gaussian to our deconvolved result. Instead, we use an ad-hoc procedure. We deconvolve Eq. (3.58) regularizing the procedure in Fourier space :

$$\widehat{p_{\log z}}(\xi) = \frac{\overline{\widehat{p_{\log|x|}}}(2\xi)\ \widehat{p_{\log|u|}}(2\xi)}{\gamma + |\widehat{p_{\log|x|}}(2\xi)|^2} \tag{3.60}$$

where $\gamma > 0$. The Fourier inverse transform of the last result gives us a first estimate of $p_{\log z}$. We keep its positive part and truncate both tails to get rid of possible oscillation artifacts leftover from the deconvolution and ensure that $E\{z\} = 1$ (i.e. $\int e^u\ p_{\log z}(u)du = 1$). We find that in the particular case of the galaxy cluster component, the prior $p_{\log z}$ is not symmetrical. It is not well fitted by a Gaussian and therefore, the prior p_z we obtain is not log-normal (see next section 3.4).

Summary

Given a template of object f^m, the procedure we follow to determine the priors $p_{z_j^m}$ is :

- If f^m is known to be well modeled by a Gaussian, we set $p_{z_j^m}(x) = \delta_{\{x=1\}}$ for each scale j.
- Otherwise, for each scale j
  1. Compute the empirical distribution p_x of the set $\{f^m_{j,q,\overline{n}}\}_{\overline{n}\in\mathbb{Z}^2, q\in[\![1,Q]\!]}$
  2. If p_x is close to Gaussian, set $p_{z_j^m}(x) = \delta_{\{x=1\}}$.
  3. If p_x is not close to Gaussian and component m does not need to be precisely estimated, set $p_{z_j^m}$ to Jeffrey's prior.
  4. If p_x is not close to Gaussian and component m needs to be precisely estimated, estimate $p_{z_j^m}$ via the ad-hoc procedure described above.

3.4 Application to astrophysical data

For our astrophysical problem, we consider four components : the Cosmic Microwave Background f^1, the clusters of galaxies f^2, the infrared point sources f^3 and the Galaxy dust f^4. The beams b^l are assumed Gaussians and the noise is white. The size of the beams b^l and level of the noise σ^l at each frequency of observation are given to us. We use the *steerable pyramid* described in detail in Section 4.3 with 4 orientations. The number of scales considered depends on the resolution of the observation. The covariance matrices of the noise neighborhoods are computed via Eq. (3.50). The covariance matrices of the component neighborhoods are estimated

from a template simulation of each component (cf. 3.3.2). The Bayes least square estimate of each component is estimated following the procedure detailed in Subsection 3.2.3. To complete the description, we need to make explicit the prior we use for each component.

Astrophysicists model the Cosmic Microwave Background f^1 by a Gaussian process, therefore we naturally set the priors $p_{z_j^1}$ to Dirac probabilities concentrated in $z = 1$, for every scale j : $p_{z_j^1}(u) = \delta_{\{u=1\}}$.

The infrared point sources f^3 are bright points. Their size is typically much smaller than the resolution of the observations, so that each pixel of the map of this component is either zero or very bright. Since point sources are isolated as well, the distribution of the wavelet coefficients of the map f^2 is mostly concentrated around zero (in large portions of the maps, there are no point sources) and has large tails (corresponding to large coefficients where the point sources are located). These distributions can not be approximated by a Gaussian for any scale j. Since the point source map is not our first focus, we use Jeffrey's prior at every scale j for the infrared point sources component : $p_{z_j^3}(u) = \frac{1}{V_{max}-V_{min}}\ \delta_{\{V_{min}\leq \log u \leq V_{max}\}}\ \frac{1}{u}$, for all j.

The galactic f^3 dust is a smooth and very slow varying signal, we find that it is reasonable to approximate the distribution of its wavelet coefficients at every scale by a Gaussian. Therefore, we set $p_{z_j^4}(u) = \delta_{\{u=1\}}$.

Finally the galaxy cluster component f^2 is the component that we want to reconstruct most accurately. The clusters of galaxies are compact objects scattered in the sky, and consequently (same reasoning as for the point sources) the distribution of their wavelet coefficients for each scale is not well approximated by a Gaussian. In order to obtain preliminary results for the reconstruction of the clusters of galaxies we will use Jeffrey's prior. In an attempt to make a better estimation, we use the ad-hoc procedure of subsection 3.3.3 to derive an improved prior for the clusters of galaxies. We display the obtained prior $p_{\log z}$, that we will refer to as the *profile*, in the top panel of Figure 3.3 with the dashed line. The Gaussian prior is plotted in plain and the log-uniform (i.e. Jeffrey's) prior is the dash dotted line. The bottom left (resp. right) panel of the figure shows the (resp. logarithm of the) marginal distribution p_x, where x is the corresponding one-dimensional Gaussian scale mixture. The pluses indicate the experimental data used to estimate the profile. As one can see on the bottom left panel, both Jeffrey's prior and our profile tend to overestimate the probability distribution around $|x| = 0$. As a consequence the number of low intensity clusters and their intensity will tend to be underestimated in the maps reconstructed using these priors. To remedy this effect, we further truncate the profile we obtained to diminish the weight of small values of $\log z$. The result is called *truncated profile* and is displayed in the Figure 3.3 by a dashed and stars curve.

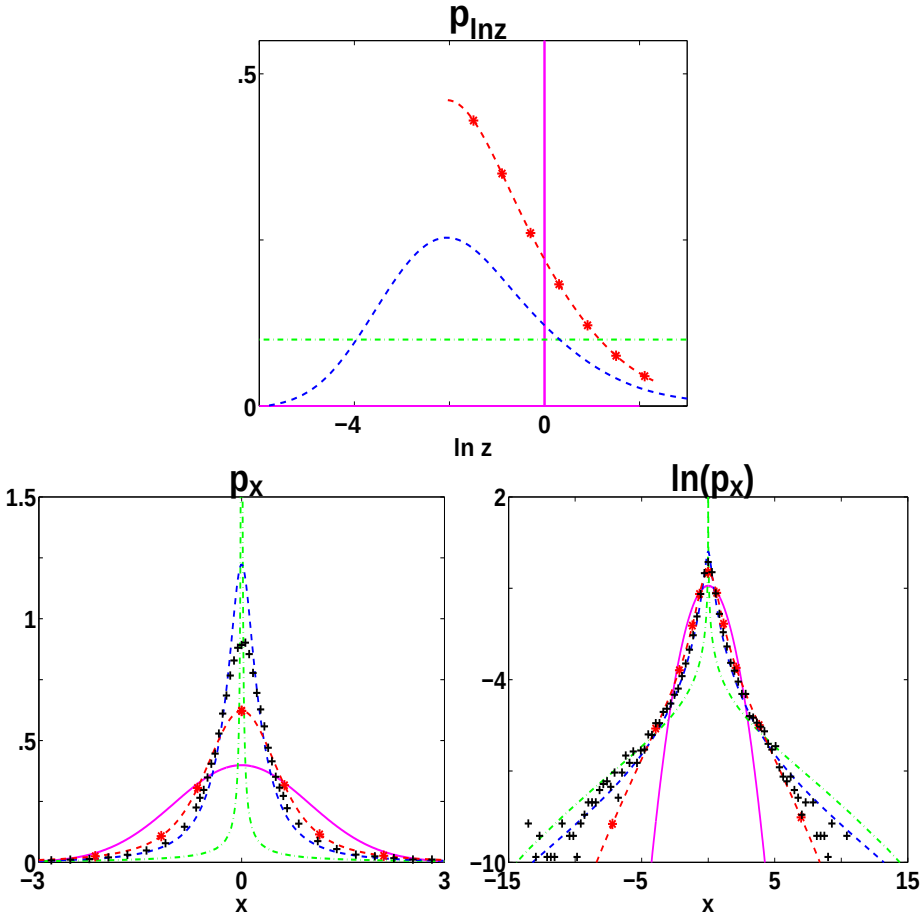

FIG. 3.3 – Top : the prior distribution of the logarithm of the multiplier $p_{\log z}$. Bottom left : p_x, the distribution of $x \equiv \sqrt{z}u$. Bottom right, the logarithm of this distribution : $\ln(p_x)$. Plain : Gaussian prior, x is Gaussian. Dash-dot : p_x corresponding to the Jeffrey's non-informative prior. Dashed : p_x corresponding to the *profile* computed from the data. Dashed and stars : p_x corresponding to the *truncated profile* computed from the data. Plus : experimental distribution p_x.

Chapitre 4

Redundant wavelet transforms

In this chapter, we review the transformations that we have utilized to decompose the signals. As we argued several times in Chapter 2 and Chapter 3, wavelet transformations have properties that we can exploit in both algorithms in order to make better estimates of the signals. The following properties are of particular interest to us : the wavelet transforms of the signals we would like to estimate are rather sparse whereas the wavelet transform of the noise is spread out ; the joint statistics of the wavelet coefficients of the components we would like to extract are well modeled by Gaussian Scale Mixtures ; some particularly useful functional vector spaces can be characterized by norms computed in wavelet space.

These properties are true for any reasonable wavelet transform. Hence, one could use any of them interchangeably without altering the arguments we presented in Chapter 2 and Chapter 3. In this chapter, we wish to give more details about two transforms that we chose to use : the *steerable pyramid* was used for the statistical algorithm presented in Chapter 3 ; the *dual tree complex wavelet transform* was used for the variational functional algorithm presented in Chapter 2.

Both of these are redundant wavelet transforms. (In a redundant transform, the generating elements can be linearly dependent). Using redundant systems, also called frames, is usually computationally more intensive and sometimes technically more difficult (e.g. subsection 2.2.3) than using bases. However there are several advantages to do so. Orthonormal wavelet transforms are not translation-invariant (because of the decimation at each scale, the wavelet transform of a translated signal is generally not the translated version of the wavelet transform of the original signal). This lack of invariance by translation is known to cause artifacts in signal processing [39, 25]. To overcome this problem, it has been proposed to use the undecimated wavelet transform, which amounts to using all possible translated wavelet bases in conjunction. The undecimated wavelet transform is redundant and computationally more intensive than the orthonormal wavelet transform. But it is translation-invariant and its use improves the quality of the processed signals [12, 35, 25]. Another drawback of the critically sampled wavelet bases is the lack of invariance by rotation ; this too can be overcome by using redundant transform. Separable wavelet bases have preferred directions along the natural axis and diagonals (in two-dimensions, horizontal, vertical and diagonal). Allowing the generating family to be redundant makes it possible to

design a frame that is tuned to more directions. For example, the steerable pyramid can be designed to be selective to any number of directions [56]. Widening the direction selectivity is of course only an approximation to rotation invariance, however it has proved to be beneficial in general image processing problems. Rotation invariance will be useful to study in detail the shape of the clusters of galaxies and the structure surrounding them, since these are highly asymmetrical objects. Finally, for the application to astrophysical data, it is quite useful to be able to characterize the power spectrum of the signals in hand in wavelet space. Indeed, the power spectrum is a quantity well-studied by astrophysicists and therefore, it can be used to incorporate a priori knowledge on the signals. The rectangular frequency tiling of orthonormal wavelet transforms does not lend itself easily to the incorporation of knowledge on the power spectrum of a signal. Once again, more flexibility is given by relaxing the linear independence condition : redundant systems can be designed to have a spherical frequency tiling (as is the case of the steerable pyramid), or to approximate it better than standard wavelet bases (as is the case of the complex wavelet transform).

In order the facilitate the presentation of the complex wavelet transform in Section 4.2 and of the steerable pyramid in section 4.3, we first review rapidly the standard orthonormal wavelet transform in section 4.1.

4.1 Orthonormal wavelet bases

Although there are other ways to define wavelet bases, we will start here from multiresolution spaces as in [40]. Subsequently, we define the scaling function ϕ and wavelet ψ, as well as the spatial filters h and g and their Fourier transform the conjugate mirror filters m_o and m_1. (One could actually start from the filters and scaling function to define the wavelet.)

4.1.1 Multiresolution analysis

Definition 4.1.1. *A multiresolution analysis of $L^2(\mathbb{R})$ is a sequence of approximation vector spaces : $\{V_j\}_{j\in\mathbb{Z}}$ that have the following properties :*

P1. $\cdots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \cdots$

P2. $\overline{\bigcup\limits_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$

P3. $\bigcap\limits_{j\in\mathbb{Z}} V_j = \{0\}$

P4. $f \in V_j \Leftrightarrow f(2^j\,\cdot) \in V_0$

P5. $f \in V_0 \Rightarrow f(\,\cdot - n) \in V_0,\ \forall n \in \mathbb{Z}$

P6. There exists ϕ in V_0 such that $\{\phi(\cdot - n), n \in \mathbb{Z}\}$ in an orthonormal basis of V_0.

The function ϕ is called the *scaling function*. Properties P4 and P6 imply that for any j, the family $\{\phi_{j,n}(\cdot) = 2^{\frac{j}{2}}\phi(2^j\cdot - n), n \in \mathbb{Z}\}$ is an orthonormal basis of V_j. Noting $h_n = \langle \phi_{-1,0}\ ,\ \phi_{0,n}\rangle$, and $m_o(\xi) = 2^{-\frac{1}{2}}\sum_{n\in Z} h_n e^{-in\xi}$, properties P1 and P6 imply :

$$\phi\big(\tfrac{x}{2}\big) = \sqrt{2}\ \sum_{n\in\mathbb{Z}}\ h_n\ \phi(x-n) \tag{4.1}$$

$$\widehat{\phi}(\xi) = m_o\big(\tfrac{\xi}{2}\big)\widehat{\phi}\big(\tfrac{\xi}{2}\big) \tag{4.2}$$

The *wavelet* can be then defined as the function ψ such that :

$$\psi\big(\tfrac{x}{2}\big) = \sqrt{2}\ \sum_{n\in\mathbb{Z}}\ g_n\ \phi(x-n) \tag{4.3}$$

$$\widehat{\psi}(\xi) = m_1\big(\tfrac{\xi}{2}\big)\widehat{\phi}\big(\tfrac{\xi}{2}\big) \tag{4.4}$$

with

$$g_n = (-1)^{1-n}\bar{h}_{1-n} \tag{4.5}$$

$$m_1\big(\xi\big) = 2^{-\frac{1}{2}}\sum_{n\in Z} g_n e^{-in\xi} = e^{-i\xi}m_o\big(\pi-\xi\big) \tag{4.6}$$

In that case, the vector space spanned by the family of translated versions of ψ, $W_0 = \text{span}\{\psi_{0,n}(\cdot) = \psi(\cdot - n),\ n \in \mathbb{Z}\}$, is the orthogonal supplement of V_0 in V_1 : $V_1 = V_0 \overset{\perp}{\oplus} W_0$. Moreover the $\{\psi_{0,n},\ n \in \mathbb{Z}\}$ are orthogonal to each other. Each approximation space V_j of the multiresolution analysis is then similarly decomposed into an orthogonal sum : $V_j = V_{j-1}\overset{\perp}{\oplus} W_{j-1}$, where V_{j-1} is the next coarser approximation space after V_j and $W_{j-1} = \text{span}\{\psi(2^{j-1}\ \cdot\ -n),\ n \in \mathbb{Z}\}$ is a detail space. It follows that the W_j span $L^2(\mathbb{R})$ and are orthogonal to each other. One can therefore consider different decompositions of $L^2(\mathbb{R})$, either using only the detail spaces W_j : $L^2(\mathbb{R}) = \overset{\perp}{\underset{j\in\mathbb{Z}}{\oplus}} W_j$ (4.7), or stopping the refinement at a particular scale J_o : $L^2(\mathbb{R}) = V_{J_o} \overset{\perp}{\underset{j\geq J_o}{\oplus}} W_j$ (4.8). The corresponding orthonormal bases are :

$$\{\ \psi_{j,n}(\cdot) = 2^{-\frac{j}{2}}\ \psi(2^{-j}\ \cdot\ -n)\ \}_{(j,n)\in\mathbb{Z}^2} \tag{4.7}$$

and

$$\{\ \phi_{J_o,n}(\cdot) = 2^{-\frac{J_o}{2}}\phi(2^{-J_o}\ \cdot\ -n)\ \}_{n\in\mathbb{Z}}\ \cup\ \{\ \psi_{j,n}(\cdot) = 2^{-\frac{j}{2}}\psi(2^{-j}\ \cdot\ -n)\ \}_{j\geq J_o,n\in\mathbb{Z}} \tag{4.8}$$

Note that the scaling function, the wavelet, the filters h and g and the conjugate filters m_o and m_1 inherit special properties from the multiresolution setting. For example, the conjugate filter m_o verifies :

$$|m_o(\xi)|^2 + |m_o(\xi+\pi)|^2 = 1\ a.e; \tag{4.9}$$

and the scaling function integrates to 1 whereas the wavelet integrates to 0.

The properties of the wavelet can be studied and adjusted by looking at the filters. The wavelets $\psi_{j,n}$ and scaling functions $\phi_{j,n}$ can have many properties that can be tailored to the application at hand, by adjusting the filter choice. For instance, one can choose to emphasize their smoothness, their localization in space and/or in frequency and the number of vanishing moments of ψ. Typically, one cannot optimize all of these simultaneously and some trade-offs have to be made. See e.g. [39, 17].

4.1.2 Computing the wavelet transform in one dimension

For a function f in $L^2(\mathbb{R})$, the wavelet decomposition corresponding to (4.7) reads :

$$f = \sum_{j\in\mathbb{Z}} \sum_{n\in\mathbb{Z}} \langle\ f\ ,\ \psi_{j,n}\ \rangle\ \psi_{j,n} \tag{4.10}$$

and alternatively, stopping the refinement at scale J_o as in (4.8) leads to :

$$f = \sum_{n\in\mathbb{Z}} \langle\ f\ ,\ \phi_{J_o,n}\ \rangle\ \phi_{J_o,n} + \sum_{j\in\mathbb{Z}} \sum_{n\in\mathbb{Z}} \langle\ f\ ,\ \psi_{j,n}\ \rangle\ \psi_{j,n} \tag{4.11}$$

The relations (4.1), (4.2), (4.3) and (4.4) propagate to the scaling coefficients $a_{j,n} = \langle\ f\ ,\ \phi_{j,n}\ \rangle$ and to the wavelet coefficients $d_{j,n} = \langle\ f\ ,\ \psi_{j,n}\ \rangle$. Indeed these coefficients can be rewritten :

$$\begin{aligned}
a_{j,n} &= \langle\ f\ ,\ \phi_{j,n}\ \rangle && (4.12)\\
a_{j,n} &= \langle\ f\ ,\ 2^{-\frac{j}{2}}\ \phi\big(2^j \cdot - n\big)\ \rangle && (4.13)\\
a_{j,n} &= \langle\ f\ ,\ 2^{-\frac{j}{2}}\ \phi\big(2^j\ (\cdot - 2^{-j}n)\ \big)\ \rangle && (4.14)\\
a_{j,n} &= \langle\ f\ ,\ \ \phi_{j,0}(\cdot - 2^{-j}n\ )\ \rangle && (4.15)\\
a_{j,n} &= \big(f * \widetilde{\phi_{j,0}}\big)(2^{-j}n) && (4.16)
\end{aligned}$$

and similarly :

$$d_{j,n} = \big(f * \widetilde{\psi_{j,0}}\big)(2^{-j}n) \tag{4.17}$$

Here, $*$ denotes the convolution on the real line and $\widetilde{\psi}(x) = \overline{\psi}(-x)$.

Fast wavelet transform in space

Using Eq. (4.1), (4.3), (4.16) and (4.17) gives formulas to compute the scaling coefficients $a_{j,n}$ and the wavelet coefficients $d_{j,n}$ from solely the scaling coefficients at the finer scale $j+1$ and the filters h and g :

$$\begin{aligned}
a_{j,n} &= (a_{j+1,\,.} \star \bar{\bar{h}}(2n) && (4.18)\\
d_{j,n} &= (a_{j+1,\,.} \star \bar{\bar{g}}(2n) && (4.19)
\end{aligned}$$

Here, $\star$ denotes the discrete convolution and $\bar{\bar{h}}_n = \bar{h}_{-n}$, $\bar{\bar{g}}_n = \bar{g}_{-n}$ This means that to find the wavelet (resp. scaling) coefficients at scale j, one computes the convolution of the scaling coefficients at scale $j+1$ with the filter $\bar{\bar{g}}$ (resp. $\bar{\bar{h}}$) and keep only the even entries.

The inverse operation : synthesizing the scaling coefficients at scale $j+1$ from the wavelet and scaling coefficients at scale j is just as simple :

$$a_{j+1,n} = (\tilde{a}_{j,\,.} \star h)(n) + (\tilde{d}_{j,\,.} \star g)(n) \tag{4.20}$$

Here, $\tilde{a}_{j,2p} = a_{j,p}$ and $\tilde{a}_{j,2p+1} = 0$ (and similarly for $\tilde{d}$). The wavelet (resp. scaling) coefficients at scale $j+1$ are interleaved with zeros and the result is convolved with

the filter h (resp. g). The sequence of scaling coefficients at scale $j+1$ is then the sum of these two convolutions.

Starting from scaling coefficients at a fine scale J_1, $\{a_{J_1,n}\}_{n\in\mathbb{Z}}$, one can recursively compute the wavelet and scaling coefficients for all scale $J_o \leq j < J_1$ using Eq. (4.18) and (4.19), for any arbitrary $J_o < J_1$. Keeping only the wavelet coefficients $d_{j,n}$ for all scales $J_o \leq j < J_1$ and the scaling coefficients $a_{J_o,n}$ at the coarsest scale , one can reconstruct the sequences of scaling coefficients at each scale from J_o to J_1 using Eq. (4.20).

The forward and inverse transform are both fast to compute since they involve only discrete convolutions and downsampling (dropping the even entries in Eq.(4.18) and (4.19)) or upsampling (adding zeros in Eq. (4.20)) two sequences at a each scale.

Wavelet transform in the frequency plane

One can rewrite Eq.(4.18) and (4.19) using the conjugate filters m_o and m_1 :

$$\widehat{a_{j,.}}(\xi) = \widehat{a_{j+1,.}}(\tfrac{\xi}{2})\,\overline{m_o}(\tfrac{\xi}{2}) \quad (4.21)$$

$$\widehat{d_{j,.}}(\xi) = \widehat{a_{j+1,.}}(\tfrac{\xi}{2})\,\overline{m_1}(\tfrac{\xi}{2}) \quad (4.22)$$

Here, for a sequence $\{v_n\}_{n\in\mathbb{Z}}$, $\widehat{v}$ denotes the trigonometric series $\widehat{v}(\xi) = \sum_{n\in\mathbb{Z}} v_n e^{-in\xi}$. From the trigonometric series, one can recover v : $v_n = \check{\widehat{v}}_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} \widehat{v}(\xi) e^{in\xi}$. To compute the wavelet (resp. scaling) coefficients at scale j with this method, one first calculates the trigonometric series associated with the scaling coefficients at scale $j+1$, then multiplies it by the conjugate filter m_1 (resp. m_o) and finally dilates the result by a factor 2. The coefficients at scale j are the Fourier coefficients of the series obtained. Note that the downsampling is done automatically here by inverting the dilated trigonometric series.

Similarly the inverse transform that computes the scaling coefficients at scale $j+1$ from scaling and wavelet coefficients at scale j can be done in Fourier space by noticing that :

$$\widehat{a_{j+1,.}}(\xi) = \widehat{a_{j,.}}(2\xi)\, m_o(\xi) + \widehat{d_{j,.}}(2\xi)\, m_1(\xi) \quad (4.23)$$

This method is not as fast as the spatial method to compute wavelet transform in the case where the spatial filters h and g have finite length, i.e. when the wavelet have compact support in space. However, in the event where the design of the wavelets has been done in the frequency plane, e.g. when the wavelet have compact support in frequency, then the spatial filters h and g are infinite and the convolution are easier to handle by this method.

The complex wavelet transform that we review in the next section is computed using spatial filters as in the fast wavelet transform, whereas the steerable pyramid transform is computed in the frequency plane.

Practical implementation with discrete signals

In practice, one has access only to a finite number of regular samples of the function f at a finite and possibly very fine scale. One considers these samples $\{f_n\}_{n\in I}$

to be the scaling coefficients at the finer scale J_1 : $f_n = \langle\ f\ ,\ \phi_{J_1,n}\ \rangle,\ n \in I$. The wavelet transform is computed neither at finer scales than J_1, nor at very coarse scales ($j \to -\infty$), where the extent of the scaling function would be greater than the support of the sample in hand. Hence, in practice, a coarse scale J_o and a fine scale $J_1 > J_o$ are naturally defined by the signal in hand.

The wavelet transform consists in the wavelet coefficients at each scale j from J_1 down to J_o, i.e. the $\{d_{j,n}\}_{J_o \le j < J_1,\ n \in I_j}$, and the scaling coefficients at the coarsest scale J_o, i.e. the $\{a_{J_o,n}\}_{n \in I_{J_o}}$. Because of the downsampling in Eq.(4.18) and (4.19), the cardinality of I_j is $|I_j| = |I|\ 2^{j-J_1}$. Note that the number of wavelet and scaling coefficients in the transform is exactly the same as the initial number of samples. This was bound to happen since the wavelet transform presented here is nothing more than a change a orthonormal basis in a finite dimensional space.

4.1.3 Separable wavelet transform in higher dimensions

In two or more dimensions, orthonormal wavelet bases are defined by taking the tensor product of several one-dimensional multiresolution analysis. Let us explain the two-dimensional case since in higher dimensions, the procedure generalizes without problems.

Definition 4.1.2. *From a multiresolution analysis of $L^2(\mathbb{R})$ $\{V_j\}_{j\in\mathbb{Z}}$ as defined in 4.1.1, the following tensor product $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ defined by :*

1. $\mathbf{V}_o = V_o \otimes V_o = \{F(x_1, x_2) = f(x_1)g(x_2), (f, g) \in V_o^2\}$
2. $F \in \mathbf{V}_j \Leftrightarrow F(2^j \cdot, 2^j \cdot) \in \mathbf{V}_o$

defines multiresolution analysis in $L^2(\mathbb{R}^2)$, i.e. $V_j \subset V_{j+1}$, $\overline{\cup_j V_j} = L^2(\mathbb{R}^2)$ and $\cap_j V_j = \{0\}$.

The approximation space $\mathbf{V}_{j+1}$ is then naturally refined into one coarser approximation space $\mathbf{V}_j = V_j \otimes V_j$ and three detail spaces : $\mathbf{W}_j^1 = V_j \otimes W_j$, $\mathbf{W}_j^2 = W_j \otimes V_j$ and $\mathbf{W}_j^3 = W_j \otimes W_j$. The corresponding orthonormal bases are :

- for $\mathbf{V}_{j+1}$: $\{\ \phi_{j,n_1}(x_1)\phi_{j,n_2}(x_2)\}_{(n_1,n_2)\in\mathbb{Z}^2}$
- for $\mathbf{W}_{j+1}^1$: $\{\ \phi_{j,n_1}(x_1)\psi_{j,n_2}(x_2)\}_{(n_1,n_2)\in\mathbb{Z}^2}$
- for $\mathbf{W}_{j+1}^2$: $\{\ \psi_{j,n_1}(x_1)\phi_{j,n_2}(x_2)\}_{(n_1,n_2)\in\mathbb{Z}^2}$
- for $\mathbf{W}_{j+1}^3$: $\{\ \psi_{j,n_1}(x_1)\psi_{j,n_2}(x_2)\}_{(n_1,n_2)\in\mathbb{Z}^2}$

Therefore the orthonormal basis considered for $L^2(\mathbb{R}^2)$ is :

$$\{\phi_{j,n_1}(x_1)\phi_{j,n_2}(x_2),\ \phi_{j,n_1}(x_1)\psi_{j,n_2}(x_2),\ \psi_{j,n_1}(x_1)\phi_{j,n_2}(x_2),\ \psi_{j,n_1}(x_1)\psi_{j,n_2}(x_2)\}_{(j,n_1,n_2)\in\mathbb{Z}^3} \tag{4.24}$$

Note that this is different from taking the tensor product of the one-dimensional wavelet basis (which would include terms mixing scales : $\psi_{j_1,n_1}(x_1)\psi_{j_2,n_2}(x_2)$).

Define d^i_{j,n_1,n_2} as the wavelet coefficients corresponding to $\mathbf{W}_j^i$ and a_{j,n_1,n_2} the scaling coefficients. The two-dimensional orthonormal wavelet transform then inherits a fast algorithm using the spatial filters h and g successively in each direction x_1 and

x_2 :

$$a_{j,n_1,n_2} = \left((a_{j+1,\cdot,\cdot} \overset{x_1}{\star} \bar{\bar{h}}) \overset{x_2}{\star} \bar{\bar{h}} \right) (2n_1, 2n_2) \quad (4.25)$$

$$d^1_{j,n_1,n_2} = \left((a_{j+1,\cdot,\cdot} \overset{x_1}{\star} \bar{\bar{h}}) \overset{x_2}{\star} \bar{\bar{g}} \right) (2n_1, 2n_2) \quad (4.26)$$

$$d^2_{j,n_1,n_2} = \left((a_{j+1,\cdot,\cdot} \overset{x_1}{\star} \bar{\bar{g}}) \overset{x_2}{\star} \bar{\bar{h}} \right) (2n_1, 2n_2) \quad (4.27)$$

$$d^3_{j,n_1,n_2} = \left((a_{j+1,\cdot,\cdot} \overset{x_1}{\star} \bar{\bar{g}}) \overset{x_2}{\star} \bar{\bar{g}} \right) (2n_1, 2n_2) \quad (4.28)$$

Here, $\overset{x_1}{\star}$ denotes the one-dimensional convolution in the direction x_1 computed for each value of n_2 (and vice-versa for $\overset{x_2}{\star}$).

As previously in one dimension, one can also consider doing these computations in the frequency plane using the conjugate filters $\overline{m}_o$ and $\overline{m}_1$ successively for x_1 and x_2. The inverse transform is also computed successively in each direction, using the spatial filters h and g and the complex conjugate filters m_o and m_1 as in Eq.(4.20) and (4.23).

The two-dimensional separable orthonormal basis presented here is sensitive to three principal directions corresponding to the detail spaces $\mathbf{W}^1$, $\mathbf{W}^2$ and $\mathbf{W}^3$: the horizontal, the vertical and the diagonal respectively. To remedy this, the complex wavelet transform combines several separable orthonormal bases that have special relations together whereas the steerable pyramid is based on the definition of radial (hence non separable) filters.

4.1.4 Other wavelet bases

Before we turn to these redundant systems, let us mention that there exist other wavelet families that are not necessarily orthonormal but still form bases of $L^2(\mathbb{R})$.

The biorthogonal wavelets can be designed to be symmetric with compact support [11]. Such a family $\{\psi^1_{j,n}\}_{j,n}$ can not form an orthonormal basis. Instead $\{\psi^1_{j,n}\}_{j,n}$ is a Riesz basis of $L^2(\mathbb{R})$ and is associated with a dual family $\{\psi^2_{j,n}\}_{j,n}$. The first wavelet is used for analysis whereas the second one is used for the reconstruction. The orthogonal relation $\langle \psi^1_{j,n},\ \psi^2_{j',n'} \rangle = \delta_{j,j'}\delta_{n,n'}$ ensures perfect reconstruction of any signal in $L^2(\mathbb{R})$.

Wavelet packets are another kind of orthonormal bases one can form starting with the same procedure as in 4.1.1. The difference is that one is allowed to further refine the vector spaces W_j by using the filter g_n and h_n on the detail coefficients d_n. (See [13, 39] for details.)

Both wavelet packets and biorthogonal wavelets can be extended to higher dimensions in a separable manner. Although they have advantages and disadvantages compared with the orthonormal wavelet transform, they share its lack of invariance by translation and poor directional selectivity. As mentioned in the introduction of this chapter, these inconveniences can be bypassed by relaxing the linear independence conditions and using frames instead of bases.

4.2 Dual tree complex wavelet transform

The complex wavelet transform has been designed originally by Kingsbury [31, 32] to remedy two principal drawbacks of traditional separable wavelet transforms in two dimensions : the lack of shift-invariance and the poor directional selectivity. The complex wavelet transform is a combination of several standard wavelet transforms, (exactly 2^n of these, where n is the dimension), that have special relations with each other. The redundancy is 2^n and the complexity is exactly 2^n times the complexity of a standard wavelet transform. This makes it just as fast to compute as a standard wavelet transform for low dimensions, in particular for images ($n = 2$).

As a consequence of the special relations between the standard transforms used in the complex transform, the latter is shift invariant in the sense that the reconstruction obtained from each scale separately is free of aliasing.

Standard wavelet coefficients oscillate rapidly close to sharp transitions. Thresholding techniques with critically sampled wavelet transforms suffer from these oscillations which cause artifacts in the reconstructions. Another advantage of the complex wavelet transform is that the modulus of the complex coefficients does not oscillate as much. Hence, the thresholding operation with complex wavelets as defined in subsection 2.2.3 causes much less artifacts.

In two dimensions, the complex wavelet transform produces 12 real wavelets. These can be paired and each pair viewed as the complex and imaginary part of a complex wavelet. In total, there are 6 complex wavelets, each one selective to a particular direction. As a consequence, the complex transform also has improved directional selectivity over standard wavelet transforms. Fig. 4.1 shows the direction selectivity achieved with the complex transform in two dimensions. The first (resp. second) row of the figure shows the 6 wavelets that can be viewed as the real (resp. imaginary) part of the 6 complex wavelet whose magnitude is shown in the last row.

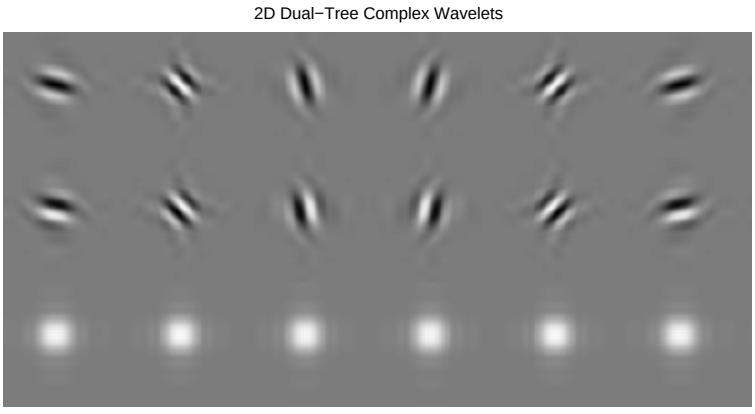

FIG. 4.1 – The complex wavelets are selective to 6 directions. First row : real part ; second row : imaginary part ; third row : amplitude of the complex wavelet. (This figure was produced by the Matlab code *cplxdual2D_plots.m* available at [66].)

4.2.1 Dual tree complex wavelet transform in one dimension

The complex wavelet transform in one dimension is implemented as two critically sampled orthonormal wavelet transforms (as described in 4.1.2) computed in parallel. Let us denote ϕ^1, ψ^1, h^1, g^1 (resp. ϕ^2, ψ^2, h^2, g^2) the scaling function, wavelet and filters relative to the first (resp. second) basis. Kingsbury in [32] showed that one way to obtain good shift invariance (as defined above) is to view the two real wavelets ψ^1 and ψ^2 as the real and imaginary part of a complex wavelet, $\Psi = \psi^1 + i\,\psi^2$, that has the property of suppressing negative frequencies :

$$\widehat{\Psi}(\xi) = 0, \text{ if } \xi < 0. \tag{4.29}$$

This happens when the two wavelets ψ^1 and ψ^2 have the special property of being Hilbert transforms of each other [52, 53], i.e. when their Fourier transform verifies :

$$\widehat{\psi^2}(\xi) = -i \text{ sign}(\xi)\, \widehat{\psi^1}(\xi), \quad \xi \in \mathbb{R} \tag{4.30}$$

This is also equivalent to designing a filter g_1 that is a half-sample delayed version of the filter g_2 :

$$g_n^2 = g_{n-\frac{1}{2}}^1 \tag{4.31}$$

Since it is not possible to design such a pair a finite impulse response filters, the Hilbert transform property has to be approximated. Selesnick [52, 53] has shown how to best do this within a preassigned filter length. It turns out that his examples correspond to those of Kingsbury [32] even though they were designed with a different criterion in mind. We shall use one of these examples implemented in the software available from Selesnick's website [66].

Suppose we have two filter banks (h^1, g^1) and (h^2, g^2), that produce wavelets ψ^1, ψ^2 that are approximate Hilbert transforms of other. The dual tree complex wavelet transform of a signal f in one dimension is computed as follows :

1. Compute the wavelet transform of f with the first filter bank (h^1, g^1) using Eq.(4.18), (4.19) to obtain the real wavelets coefficients $\{d_{j,n}^1\}_{J_o \le j < J_1,\ n \in I_j}$ and real scaling coefficients $\{a_{J_o,n}^1\}_{n \in I_{J_o}}$.
2. Compute similarly a wavelet transform of f with (h^2, g^2) to obtain a second set of real wavelets coefficients $\{d_{j,n}^2\}_{J_o \le j < J_1,\ n \in I_j}$ and real scaling coefficients $\{a_{J_o,n}^2\}_{n \in I_{J_o}}$.
3. The coefficients of the dual tree complex wavelet transform are the complex wavelet coefficients $\{c_{j,n} = d_{j,n}^1 + i\, d_{j,n}^2\}_{J_o \le j < J_1,\ n \in I_j}$, and the real scaling coefficients $\{a_{J_o,n}^1\}_{n \in I_{J_o}} \cup \{a_{J_o,n}^2\}_{n \in I_{J_o}}$.

The complex wavelet coefficients $\{c_{j,n}\}_{n \in I_{J_o}}$ can then be modified the same way one would do with real wavelet coefficients, but keeping the phase constant, as described in subsection 2.2.3. For example, soft-thresholded coefficients $\{c'_{j,n}\}_{J_o \le j < J_1,\ n \in I_j}$ would be defined the following way :

$$\text{if } \ c_{j,n} = |c_{j,n}|.e^{i\,\theta} \ \text{ then } \ c'_{j,n} = S_{\tau,1}(c_{j,n}) = \begin{cases} \big(|c_{j,n}| - \tau\big).e^{i\,\theta} & \text{if } |c_{j,n}| \ge \tau \\ 0 & \text{if } |c_{j,n}| < \tau \end{cases} \tag{4.32}$$

And one would reconstruct a signal from these by :

1. Defining the real wavelet coefficients : $d'^1_{j,n} = \Re(c'_{j,n})$ and $d'^2_{j,n} = \Im(c'_{j,n})$.
2. Reconstructing f_1 from the real scaling coefficients $\{a^1_{J_o,n}\}_{n\in I_{J_o}}$ and the real wavelet coefficients $d'^1_{j,n}$ with the filter bank (h^1, g^1) using Eq.(4.20).
3. Reconstructing f_2 from the real scaling coefficients $\{a^2_{J_o,n}\}_{n\in I_{J_o}}$ and the real wavelet coefficients $d'^2_{j,n}$ with the filter bank (h^2, g^2) using Eq.(4.20).
4. Taking the average : $\frac{f_1+f_2}{2}$.

Remark. A slight modification has to be done in practice for discrete signals. For a single real wavelet transform, we considered the samples f_n to be the scaling coefficients $f_n = \langle f, \phi_{J_1,n}\rangle$ at the finest scale. This means that the underlying function f is $f = \sum_n f_n \phi_{J_1,n}$. In the case of the dual tree complex wavelet transform, we have two different scaling functions. Considering the samples f_n as the scaling coefficients at the finest scale would mean that we are analyzing two different underlying functions : $\sum_n f_n \phi^1_{J_1,n}$ and $\sum_n f_n \phi^2_{J_1,n}$. This is clearly not the goal. Special filters have to be designed for the first stage of the transform to correct for that.

4.2.2 Dual tree complex wavelet transform in two dimensions

As we saw in the precedent section, a standard separable wavelet transform produces three wavelets : $\phi(x)\psi(y)$, $\psi(x)\phi(y)$ and $\psi(x)\psi(y)$. Again, one can compute the standard separable wavelet transform with each filter bank (h^1, g^1) and (h^2, g^2). One can define six real wavelets $\psi^{i,j}$, $i = 1, 2$, $j = 1, 2, 3$, by combining the three wavelets obtained in each transform the following way :

$$\begin{aligned}
\psi^{1,1}(x,y) &= \phi^1(x)\psi^1(y) + \phi^2(x)\psi^2(y) && (4.33)\\
\psi^{1,2}(x,y) &= \psi^1(x)\phi^1(y) + \psi^2(x)\phi^2(y) && (4.34)\\
\psi^{1,3}(x,y) &= \psi^1(x)\psi^1(y) + \psi^2(x)\psi^2(y) && (4.35)\\
\psi^{2,1}(x,y) &= \phi^1(x)\psi^1(y) - \phi^2(x)\psi^2(y) && (4.36)\\
\psi^{2,2}(x,y) &= \psi^1(x)\phi^1(y) - \psi^2(x)\phi^2(y) && (4.37)\\
\psi^{2,3}(x,y) &= \psi^1(x)\psi^1(y) - \psi^2(x)\psi^2(y)\ . && (4.38)
\end{aligned}$$

Similarly to the six wavelets displayed in the first row of Fig.4.1, each of these six wavelets is sensitive to one direction. Hence by summing and differencing the wavelets coefficients from two standard separable wavelet transforms, one gets a system of redundancy two that has good directional selectivity.

However, these six wavelets cannot be paired and considered as real and imaginary part of complex wavelets. To do so, one needs to consider two additional real separable wavelet transforms. Unlike what we described so far, these transforms do not operate the same way on rows and columns of the signal : one needs to use (h^1, g^1) to filter the rows and (h^2, g^2) to filter the columns (and conversely). By summing and differencing the outputs of the four real separable wavelet transforms, one gets the six complex

wavelets displayed in Fig.4.1. They are defined by :

$$\begin{aligned}
\Psi^{1,1}(x,y) &= \left[\phi^1(x)\psi^1(y)+\phi^2(x)\psi^2(y)\right]+i\ \left[\phi^1(x)\psi^1(y)-\phi^2(x)\psi^1(y)\right] && (4.39)\\
\Psi^{1,2}(x,y) &= \left[\psi^1(x)\phi^1(y)+\psi^2(x)\phi^2(y)\right]+i\ \left[\psi^1(x)\phi^2(y)-\psi^2(x)\phi^1(y)\right] && (4.40)\\
\Psi^{1,3}(x,y) &= \left[\psi^1(x)\psi^1(y)+\psi^2(x)\psi^2(y)\right]+i\ \left[\psi^1(x)\psi^2(y)-\psi^2(x)\psi^1(y)\right] && (4.41)\\
\Psi^{2,1}(x,y) &= \left[\phi^1(x)\psi^2(y)+\phi^2(x)\psi^1(y)\right]+i\ \left[\phi^1(x)\psi^1(y)-\phi^2(x)\psi^2(y)\right] && (4.42)\\
\Psi^{2,2}(x,y) &= \left[\psi^1(x)\phi^2(y)+\psi^2(x)\phi^1(y)\right]+i\ \left[\psi^1(x)\phi^1(y)-\psi^2(x)\phi^2(y)\right] && (4.43)\\
\Psi^{2,3}(x,y) &= \left[\psi^1(x)\psi^2(y)+\psi^2(x)\psi^1(y)\right]+i\ \left[\psi^1(x)\psi^1(y)-\psi^2(x)\psi^2(y)\right] && (4.44)
\end{aligned}$$

Remark. As in the one-dimensional case, special filters for the first stage of the transform have to be used and thresholding operations are done on the complex coefficients.

4.3 Steerable pyramid

Much like the complex wavelet transform, the steerable pyramid is a linear transformation that decomposes two-dimensional signals into subbands localized in scale and in orientation. But unlike the complex wavelet transform, this tight frame is not made of a concatenation of bases, but rather is designed from scratch by computing filters in the Fourier plane that have desired properties. One low-pass filter (like m_o), one high pass filter (like m_1) and M oriented filters that are rotated versions of a unique filter define the steerable pyramid. This corresponds to having one scaling function and M "wavelets".

The steerable pyramid transform is translation-invariant and essentially aliasing-free (the filters are designed to be band-limited so that the sampling rate is above Nyquist frequency). It can produce an arbitrary number M of orientations and therefore can approximate rotation-invariance much better than the standard separable wavelet transform. Note that, theoretically, the *steerability* of this transform makes it totally rotation-invariant : the filters are designed so that the response to any particular orientation can be computed by linear combinations of the response to the M original orientations. The steerability of the transform is the reason it was designed in the first place. However, the transform has proved to be quite efficient and useful using only the M principal orientations and that is how we shall also use it here.

4.3.1 Description of the filters, scaling functions and wavelets

In this section, we denote $\widehat{f}$ the Fourier transform of the function f and (r,θ) the polar coordinates. Moreover, we write $\overline{n}$ for the vector (n_1,n_2). As in the separable case, the scaling function is indexed by scale j and the location $\bar{n}$: $\phi_{j,\overline{n}}$. The wavelets bear an additional index m corresponding to the orientation : $\psi^m_{j,\overline{n}}$. Here, the wavelet and scaling function at the scale j are not sampled at the same rate :

$$\begin{aligned}
\phi_{j,\overline{n}}(\overline{x}) &= 2^j\ \ \phi\left(2^j\overline{x}-\overline{n}\right) && (4.45)\\
\psi^m_{j,\overline{n}}(\overline{x}) &= 2^j\ \psi^m\left(2^j\overline{x}-2\overline{n}\right) && (4.46)
\end{aligned}$$

The wavelets and scaling function verify scaling relations analogous to Eq. (4.2) and (4.4) in the separable case, with the addition of orientation for the wavelets :

$$\widehat{\phi}(2r,\theta) = \widehat{\phi}(2r) = \widehat{\phi}(r)\, L(r) \tag{4.47}$$

$$\widehat{\psi^m}(2r,\theta) = \widehat{\phi}\big(r\big)\, H\big(r\big)\, G_M(\theta - \tfrac{m\pi}{M}) \tag{4.48}$$

The low-pass filter L, the high-pass filter H and the oriented filter G_M are defined as follows :

$$L(r) = \cos\big(\tfrac{\pi}{2}\log_2(\tfrac{4r}{\pi})\big)\,\delta_{\frac{\pi}{4}<r<\frac{\pi}{2}} + \delta_{r<\frac{\pi}{4}} \tag{4.49}$$

$$H(r) = \sin\big(\tfrac{\pi}{2}\log_2(\tfrac{4r}{\pi})\big)\,\delta_{\frac{\pi}{4}<r<\frac{\pi}{2}} + \delta_{r>\frac{\pi}{2}} \tag{4.50}$$

$$G_M(\theta) = \frac{(M-1)!}{\sqrt{M[2(M-1)]!}}\,\big|2\cos\theta\big|^{M-1} \tag{4.51}$$

They are displayed for $M = 4$ in Fig.4.2.

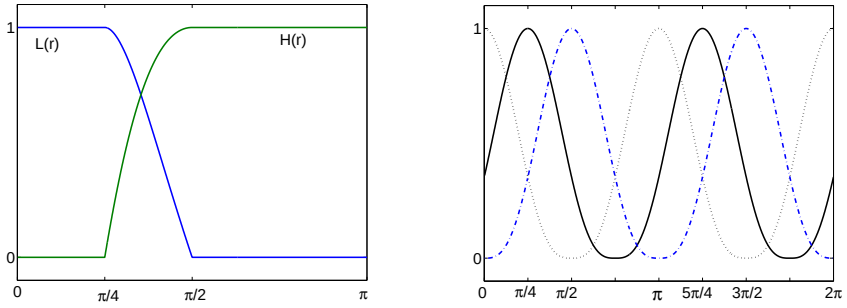

FIG. 4.2 – Left : low pass filter $L(r)$ and high-pass filter $H(r)$. Right : oriented filters $G_M(\theta - \frac{m\pi}{M})$ for $M = 4$. Dotted curve : $m = 0$, orientation of the wavelet : 0° ; plain curve : $m = 1$, orientation of the wavelet : 45° ; dash-dotted curve : $m = 2$, orientation of the wavelet : 90°. Omitted for clarity of the figure : $m = 3$, orientation of the wavelet : 135°.

The scaling function is real, non-negative and radially symmetric and so is its Fourier transform. The wavelets are real and oriented, their Fourier transform is real non-negative and symmetric about the origin. Examples of wavelets and scaling function are displayed in the first row of Fig. 4.3. The second row of the figure shows their Fourier transform. The wavelets shown have different scale, orientation and location.

Remark. We use a non-negative version of the oriented filter proposed in [47] : $G_M(\theta) = \frac{(M-1)!}{\sqrt{M[2(M-1)]!}}\,\big(2\cos\theta\big)^{M-1}$. The oriented filter we propose ensures that the wavelets are always real. It is less smooth than the original for $M = 2$, i.e. when one considers only two orientations. In that case, our G_2 is only continuous, while the one used by Portilla et al. is C^∞. However, this lack of smoothness was already present in the low-pass and high-pass filters which are continuous but not differentiable. Therefore, our choice does not change the overall regularity of the Fourier transforms of

FIG. 4.3 – Top row : wavelets in space ; Bottom row : wavelets in Fourier plane. First column : wavelet at a fine scale $j+1$, centered at location $\overline{n}_0$, oriented along the first diagonal. Second column : wavelet at a coarser scale j, centered at location $\overline{n}_1$, oriented along the first diagonal. Third column : wavelet at the same coarser scale j, centered at location $\overline{n}_2$, oriented along the horizontal axis. Fourth column : scaling function, centered at location $\overline{n}_2$.

the scaling function and wavelets even for $M = 2$. Moreover, we verified that the non-differentiability of the filters does not impact the performances of our reconstruction algorithm of astrophysical data by designing C^1 (continuously differentiable) filters :

$$L(r) = \cos\big(\tfrac{\pi}{2}\,\nu(r-1)\big) \tag{4.52}$$

$$H(r) = \sin\big(\tfrac{\pi}{2}\,\nu(r-1)\big) \tag{4.53}$$

$$G_M(\theta) = \sin\big(\tfrac{\pi}{2}\,\cos\big(\tfrac{M}{2}\theta\big)^2\big), \qquad \theta \in [-\tfrac{\pi}{M},\,\tfrac{\pi}{M}] \tag{4.54}$$

$$G_M(\theta) = \sin\big(\tfrac{\pi}{2}\,\cos\big(\tfrac{M}{2}(\theta-\pi)\big)^2\big), \quad \theta \in [\pi-\tfrac{\pi}{M},\,\pi+\tfrac{\pi}{M}] \tag{4.55}$$

$$\text{with } \nu(x) = \sin\big(\tfrac{\pi}{2}\,x\big)^2\,\delta_{0<x<1} + \delta_{x\geq 1} \tag{4.56}$$

Since the use of the C^1 filters (4.52)-(4.55) does not improve the results, we will only present our work using the filters (4.49)-(4.51).

4.3.2 Algorithm to compute the steerable pyramid transform

We use the notation : $a_{j,\overline{n}} = \langle f\,,\,\phi_{j,\overline{n}}\rangle$ for the scaling coefficients and $d^m_{j,\overline{n}} = \langle f\,,\,\psi^m_{j,\overline{n}}\rangle$ for the wavelet coefficients of a function f oriented in the direction $\frac{m\pi}{M}$. Suppose we are given the scaling coefficients at scale $j+1$: $\{a_{j+1,\overline{n}}\}_{\overline{n}\in\mathbb{Z}^2}$. The algorithm to compute the coefficients at the coarser scale in the Fourier plane is :

1. Compute the trigonometric series $\widehat{a_{j+1,\overline{n}}}(\xi)$.
2. Multiply by the high-pass filter $H(|\xi|)$, call the result $T(\xi)$.
3. For $m = 0$ to $m = M-1$, multiply T by the rotated oriented filter to obtain : $\widehat{d_{j,\overline{n}}}(\xi) = T(\xi)G_M(\theta(\xi) - \frac{m\pi}{M})$, where $\xi = |\xi|e^{i\theta(\xi)}$.
   Inverse the trigonometric series to obtain the wavelet coefficients $\{d_{j,\overline{n}}\}_{\overline{n}\in\mathbb{Z}^2}$.
4. Multiply $\widehat{a_{j+1,\overline{n}}}(\xi)$ by the low-pass filter and keep a dilated version : $\widehat{a_{j,\overline{n}}}(\xi) = \widehat{a_{j+1,\overline{n}}}(\frac{\xi}{2})L(\frac{|\xi|}{2})$

5. Inverse the last trigonometric series to find the scaling coefficients $\{a_{j,\overline{n}}\}_{\overline{n}\in\mathbb{Z}^2}$.

Given the scaling coefficients at the finest scale, it suffices to repeat this procedure recursively to find the decomposition of f on the steerable pyramid. Since the scaling coefficients are kept at only the coarsest scale, step 1 (resp. step 5) can be skipped at each iteration except the first (resp. last) one.

The reconstruction of the scaling coefficients at scale j from the wavelet and scaling coefficients at scale $j-1$ is carried out using the exact same filters L, H and G_M. Indeed, the steerable pyramid is a tight frame which ensures that the decomposition and reconstruction are done with the same family :

$$f = \sum_{\overline{n}\in\mathbb{Z}^2} \langle\, f\ ,\ \phi_{J_o,\overline{n}}\,\rangle\ \phi_{J_o,\overline{n}} + \sum_{m=1}^{M}\sum_{j\in\mathbb{Z}}\sum_{\overline{n}\in\mathbb{Z}^2} \langle\, f\ ,\ \psi^m_{j,\overline{n}}\,\rangle\ \psi^m_{j,\overline{n}}. \tag{4.57}$$

And the filters are real so that :

$$\widehat{a_{j+1,\cdot}}(\xi)\ =\ \widehat{a_{j,\cdot}}(2\xi)\ L\big(|\xi|\big)\ +\ \sum_{m=1}^{M} \widehat{d^m_{j,\cdot}}(\xi)\ L\big(|\xi|\big)\ G_M\big(\theta(\xi) - \tfrac{m\pi}{M}\big) \tag{4.58}$$

Figure 4.4 shows the system diagram corresponding to the decomposition and reconstruction. The steps described above correspond to the shaded block. In practice, the sample of the function f in hand are again considered as the scaling coefficients at the finest scale : $\{a_{J_1,\overline{n}}\}_{\overline{n}\in\mathbb{Z}^2}$. To avoid aliasing in the practical case of a finite sample, one needs to use a slightly modified version of the algorithm to compute the coefficients at scale J_1-1. As pictured in the white block of Fig. 4.4, one does not do the downsampling for the scaling coefficients at scale J_1-1, which means that :

$$\widehat{a_{J_1-1,\overline{n}}}(\xi) = \widehat{a_{J_1,\overline{n}}}(\xi) L\big(|\xi|\big). \tag{4.59}$$

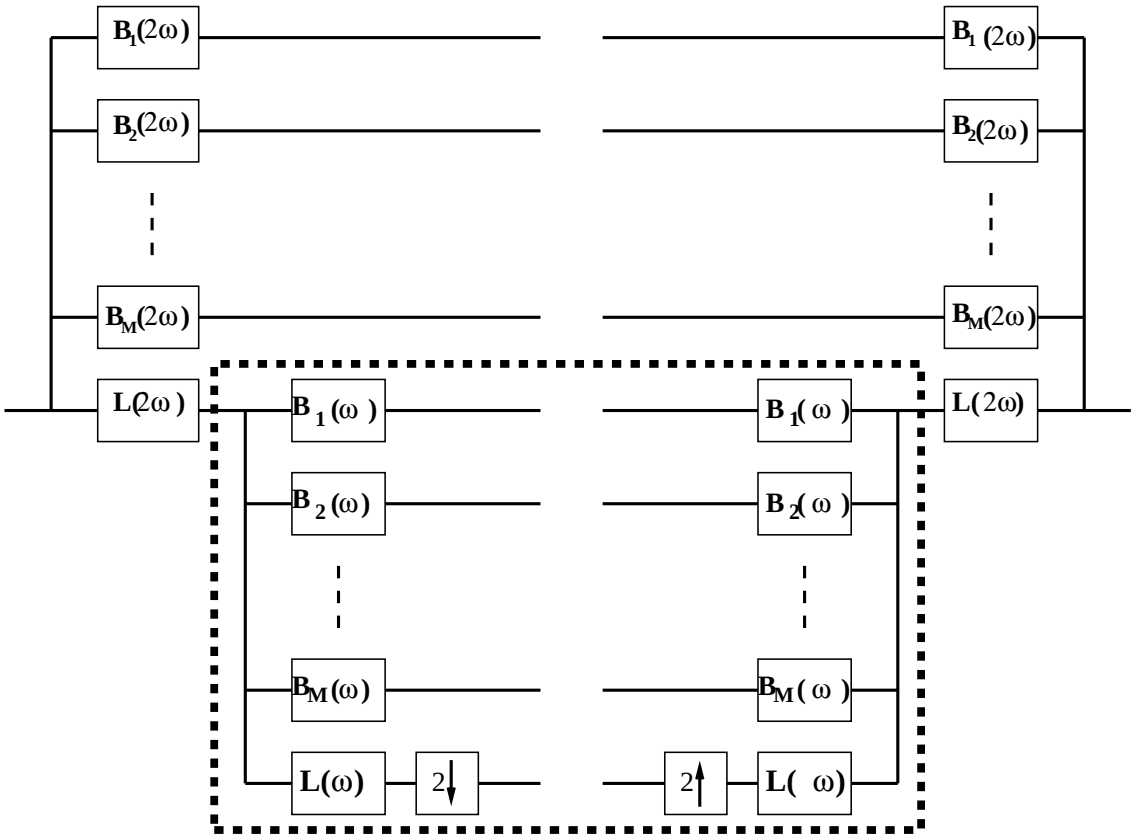

FIG. 4.4 – System diagram for the decomposition and synthesis using the steerable pyramid. The area inside the dotted line is repeated recursively to obtain the full transform. Here $\omega = (r, \theta)$, and $B_m(\omega) = H(r)G_m(\theta)$

Chapitre 5

Application to the extraction of clusters of galaxies

This chapter is dedicated to the study of the performances of the functional variational method described in Chapter 2 and the statistical method described in Chapter 3 for the reconstruction of maps of clusters of galaxies via the detection of their Sunyaev-Zeldovich signature in the fluctuations of the Cosmic Microwave Background radiation. The mathematical model that describes the observations and the components have been described in the precedent chapters. In the first section of this chapter, we explain in greater detail the cosmology of each component and show examples of simulated observations. The second section describes the tools we use to assess the quality of the reconstructed maps. In section 5.3, 5.4, and 5.5, we analyze the performance of both methods under different conditions of observation. The results we obtained are summarized in section 5.6.

5.1 Description of the signals

5.1.1 Clusters of galaxies

Stars are usually found in dense collections rather than isolated. A collection of stars (ten millions to one trillion), together with interstellar gas, dust, and dark matter, all being held together by gravitational attraction, is called a galaxy. Most galaxies are several thousand to several hundred thousand light years in diameter. Galaxies themselves are organized into larger structures. The smaller aggregates of galaxies are called groups of galaxies. Typically, a group of galaxies contains less than fifty of them. Clusters of galaxies are larger structures containing fifty to thousands of galaxies, packed into areas of around one megaparsec across (one parsec is around 3.12 light years). Superclusters are even larger structures yet, containing tens of thousands of galaxies found in groups, in clusters or even isolated. They form the largest structures identified so far in the universe, and resemble a foam.

Our work focuses on the reconstruction of clusters of galaxies because they may be used to infer cosmological information such as the Hubble constant via number

counts and power spectrum analysis of Sunyaev-Zeldovich maps (cf. [38, 36, 27, 3]). This is one of the most important scientific goals of several experiments, now planned or underway, such as the Sunyaev-Zeldovich Array experiment, the Atacama Cosmology Telescope SZ survey and the Planck mission. Galaxies in the clusters travel at velocities in the range of eight hundred to a thousand $km.s^{-1}$ and are surrounded by hot X-ray emitting gas and large amounts of dark matter. The total mass of a cluster is typically between 10^{14} and 10^{15} times the solar mass, with only five percent (resp. ten) of the mass of a cluster due to the galaxies (resp. the gas), the rest being dark matter. Reconstructing the clusters of galaxies not an easy task, because other physical phenomena, such as the Cosmic Microwave Background, obscure our view of it. However, imaging techniques have now reached a sufficient resolution that the Sunyaev-Zeldovich signature of the clusters can be extracted for further study.

The Sunyaev-Zeldovich effect (SZ effect in short) is due to high energy electrons in the galaxy clusters that interact with Cosmic Microwave Background (CMB) photons traveling from the last scattering surface to Earth. Some high energy of the electrons is transferred to the low energy photons through the inverse Compton effect. This modifies the Cosmic Microwave Background temperature and intensity in the direction of a cluster. The thermal SZ effect induces distortions of Cosmic Microwave Background spectrum, its frequency dependence is different from that of the CMB and its amplitude is comparable to the CMB fluctuations. Hence the detection of the thermal SZ signal will allow to study clusters of galaxies. The right panel of Figure 5.1 and the bottom left panel of Figure 5.2 show examples of thermal Sunyaev-Zeldovich clusters' signatures. Note that there is also a kinetic SZ effect due to the bulk motion of the clusters. This signal is much weaker than the thermal SZ signal and has a frequency dependence similar to that of the Cosmic Microwave Background, therefore we will not attempt to detect it.

5.1.2 The Cosmic Microwave Background

The Cosmic Microwave Background radiation or CMB is a form of electromagnetic radiation that fills the whole of the Universe (see Figure 5.1, left panel and Figure 5.2, top left panel, for two examples). Its existence and properties are considered one of the major confirmations of the Big Bang theory. According to standard cosmology, the CMB gives a snapshot of the Universe at the "time of last scattering", about 400,000 years after the Big Bang, when the Universe became transparent to radiation for the first time. Since this time, the Universe is expanding, causing the CMB photons to be redshifted and the radiation to cool with a factor inversely proportional to the Universe's scale length.

The CMB spectrum matches closely that of a black body at 2.726 Kelvins and this radiation has a high degree of isotropy. There are, however, anisotropies and these are the features that help us understand the Universe. The most pronounced anisotropy is the dipole anisotropy, which is consistent with the Earth moving relative to the CMB. A number of experiments, starting with the Cosmic Background Explorer (COBE) satellite in 1989-1996, have since detected large scale anisotropies other than the dipole, allowing cosmologists to understand better the structure of the Universe. For

example, the measurements were able to rule out some theories of cosmic structure formation like the cosmic strings theory. In 2000, the Boomerang experiment reported that the highest power fluctuations occur at the scale of one degree. Together with other cosmological data, these results implied that the geometry of the Universe is flat. In 2003, the WMAP experiment provided a detailed measurement of the angular power spectrum down to this scale, tightly constraining various cosmological parameters. These results are broadly consistent with those expected from cosmic inflation as well as various other competing theories.

To make further progress, it is known that smaller scale fluctuations than what was provided by WMAP will have to be analyzed. These very small scale fluctuations have been previously observed by ground-based interferometers in small regions of the sky and will be measured systematically over the whole sky by the space mission Planck, which is to be launched in the next two to three years. These small scales correspond to the scale of massive galaxy clusters (see Figure 5.1). The Sunyaev-Zeldovich signature of the clusters is a major factor of the fluctuations of the CMB at these scales. Therefore, not only will these CMB survey experiments such as Planck give data to resolve massive clusters, but also the extraction and accurate reconstruction of these clusters of galaxies will be needed to proceed with the CMB analysis.

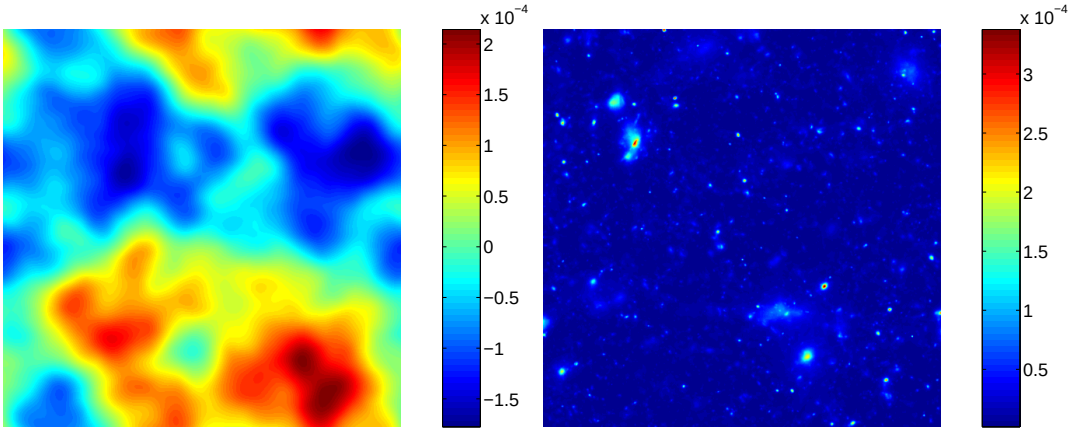

FIG. 5.1 – Simulated 1 degree by 1 degree maps. Left panel : CMB, Right panel : SZ clusters.

We consider experiments that will provide a map of the sky in the frequency range 100-600 GHz, that is, where the thermal SZ signal has the biggest amplitude. In this range, two other physical components will have a significant contribution to the observed maps : the radio and infrared point sources and the Galaxy dust. We describe briefly these two components in the next subsection.

5.1.3 Point sources and the Galaxy dust

The Galaxy dust refers to accumulations of gas and dust between stars in our own galaxy. These form an interstellar cloud that lies in the foreground of our observations

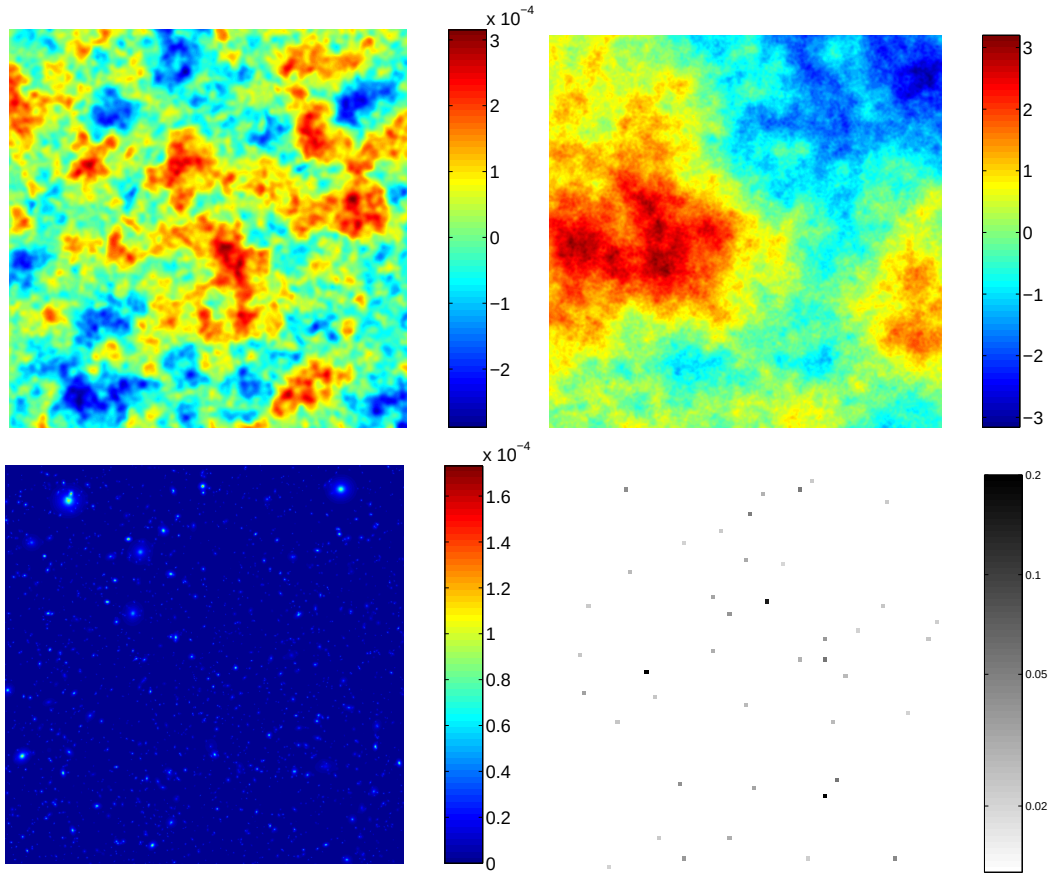

FIG. 5.2 – Simulated 10 degrees by 10 degrees maps. Top left panel : CMB, top right panel : the Galaxy dust, bottom left panel : SZ clusters, bottom right panel : infrared point sources, shown much bigger than their true size for clarity (see text). Note the difference in scale between Figure 5.1 and 5.2.

of the sky. The frequency dependence of the galactic dust is significantly different from that of the CMB and the SZ effect. Similarly to the CMB signal, the galactic dust spreads across our observations of the whole sky and its fluctuations are smooth (see Figure 5.2, top right panel for an illustration). Because the Galaxy dust has very different spatial properties from the SZ signal, we do not expect that its contributions will limit our reconstruction of the SZ clusters even though they are more faint.

On the other hand, point sources may reveal themselves to be more serious pollutants of our SZ reconstructions. Technically, the term point source could refer to any source that can be treated as coming from a single point. Here, point sources are of two types : radio galaxies, brightest in the lowest frequency channel, and dusty galaxies, brightest in the highest frequency channel. The radio point source signal is very weak in the range of frequencies we analyze and will not be considered here. Dusty star-forming galaxies at high-redshift shine brightly at submillimeter frequency, and therefore, will be a potential concern. We show at the bottom right of Figure 5.2 an

example of a simulated map of infrared point sources, each point source being extended to several pixels to allow visualization. The modeling of these infrared sources, (number counts, frequency dependences, and spatial correlations) remains uncertain. Therefore we will first concentrate our efforts on lower frequencies where the point sources can be ignored to assess the ability of our algorithms to separate the SZ effect from the CMB variations. Two analyzes, one at higher resolution and the second one at lower resolution, are made ignoring the point sources and the Galaxy dust. We incorporate these two components in a third study to complete our analysis.

5.1.4 Frequency dependences

In this section, we describe in more detail how the contribution of each astrophysical component varies with the frequency of observation. The thermal Sunyaev-Zeldovich effect causes a change in the CMB temperature in the direction $\vec{n}$:

$$\frac{\delta T_{CMB}}{T_{CMB}} = -2\ y(\vec{n})\left[2 - \frac{x}{2}\frac{\exp(x)+1}{\exp(x)-1}\right] \tag{5.1}$$

with

$$x = \frac{h\nu}{k_B T_{CMB}} \tag{5.2}$$

where ν is the frequency of observation in GHz, $h \simeq 6.626 \times 10^{-34} m^2\ kg\ s^{-1}$ is the Planck constant, $k_B \simeq 1.38 \times 10^{-23} m^2 kg\ s^{-2} K^{-1}$ is the Boltzmann constant and $T_{CMB} \simeq 2.726K$ is the CMB temperature. The comptonization parameter $y(\vec{n})$ is the quantity intrinsic to the cluster while the rest of Eq. (5.1) models the frequency dependence, when the observation is measured in CMB temperature units; that is, when the observations are normalized so that the frequency dependence of the CMB is flat. The left panel of Figure 5.3 displays the frequency dependence of the SZ signal in CMB temperature units (black or dotted line). For reference, the blue or dash dotted line is the flat frequency dependence of the CMB and it is equal to one in these units. The thermal SZ effect causes a decrement of the temperature below the characteristic frequency of 217 GHz, and an increment of the temperature above it. The effect is illustrated in the first three panels of Figure 5.4, where the location of a particular cluster is pointed by an arrow labeled with the letter "c" in three observations at different frequencies. In the top left panel, the presence of the clusters decreases the intensity measured at 145 GHz. This effect disappears in the top right panel because at 217 GHz, the frequency dependence of the SZ signal is close to zero. Finally at 265 GHz (middle left panel), the effect is inverted, the presence of the cluster causing an increase of intensity.

In CMB units, it seems that the larger the frequency of observation, the more important the SZ contribution is. However this is relative to the CMB frequency dependence itself. In fact, the SZ signal is maximal (resp. minimal) around 350 (resp. 145) GHz (see right panel of Figure 5.3), when the observations are measured in intensity units. The CMB signal itself reaches its maximum around 217 GHz, where

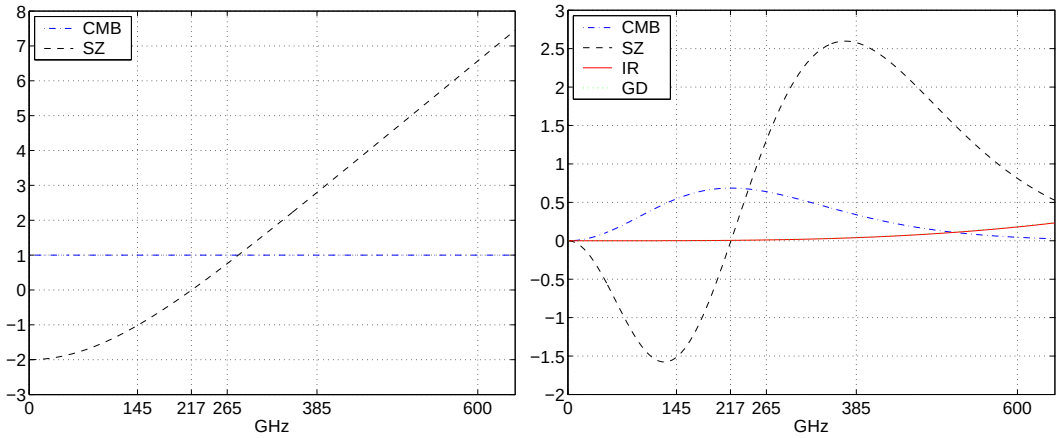

FIG. 5.3 – Frequency dependence, left panel in CMB temperature unit, right panel in flux units. Note that the frequency dependence of point sources (IR) and Galaxy dust (GD) in the right panel coincide.

the clusters' dependence changes signs (Figure 5.3, right panel, blue dash dotted line). The plain and dotted curves displaying the frequency dependence of the infrared point sources and the Galaxy dust lay on top of each other in this figure.

To obtain a complete picture of the contribution of each component to the observation at each frequency, one should bear in mind that the natural units of each components are different. The frequency dependences displayed in Figure 5.3 take these units into account. For example, the CMB signal is measured in Kelvin, which is the unit used in the left panel (top left panel) of Figure 5.1 (5.2). Its fluctuations are of the order of 10^{-4} Kelvin. The SZ cluster signal is measured by its comptonization parameter y, also called y-parameter. The order of magnitude of the y-parameter of the most massive and brightest clusters is around 10^{-4} as well (right and top right panels of Figure 5.1 and 5.2). Combining this with the frequency dependences, one can see that massive clusters yield a signal of amplitude that is comparable to that of the CMB in the range of frequencies observed. This is not the case for the point sources signal and the Galaxy dust signal. The natural unit for these signals is the flux at a particular frequency and although their frequency dependence stays below the SZ frequency dependence (see Figure 5.3, right panel), those two signals are the dominant signals at higher frequencies.

Figure 5.4 gives a visual summary of these remarks. Each panel shows a 3.2 by 3.2 degrees maps containing the sum of the contributions of the four signals at a particular frequency. This result is convolved with a two arcminutes wide beam so that the contribution of the points sources is wide enough to be visible, without the artificial blowing up used in Figure 5.2. The middle right panel and bottom panels show that above 300 GHz, the point sources and the Galaxy dust are dominating the CMB and SZ signals. In the 100-300 GHz range on the other hand, the CMB signal is dominant and traces of SZ clusters can be seen, as pointed out by the arrow labeled with the letter "c". This suggests that the relevant frequencies of observation for the extraction and detection of the SZ clusters' signal are between one hundred and three hundred

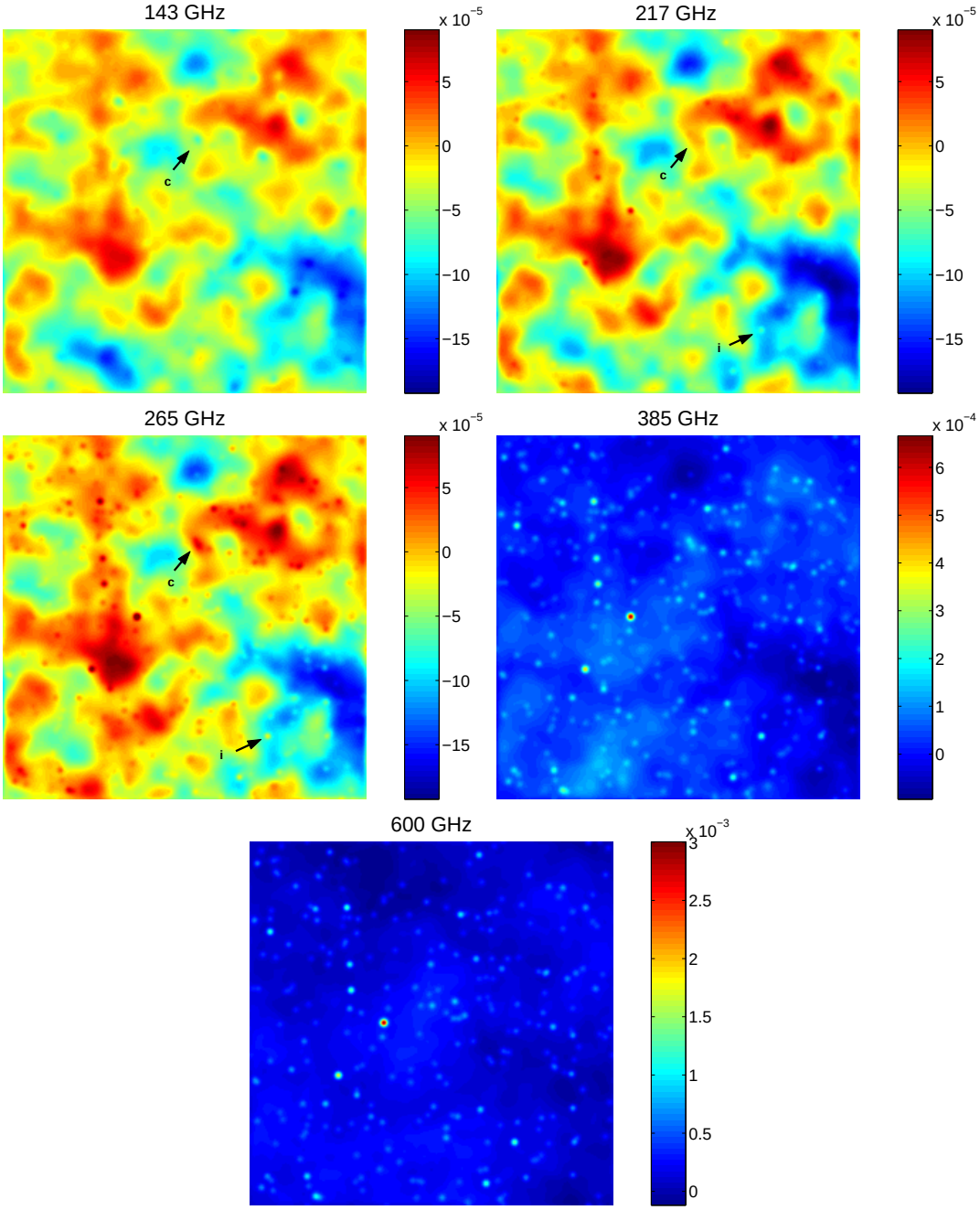

FIG. 5.4 – Simulated 3.2 by 3.2 degrees maps of the sum of the contribution of the CMB, the thermal SZ, the infrared point sources and the Galaxy dust at different frequencies of observation. Top left : 143 GHz, top right : 217 GHz, middle left : 265 GHz, middle right : 385 GHz, bottom : 600 GHz. One particular cluster of galaxies is located by the arrow labeled with "c". One particular infrared point source is located by the arrow labeled with "i".

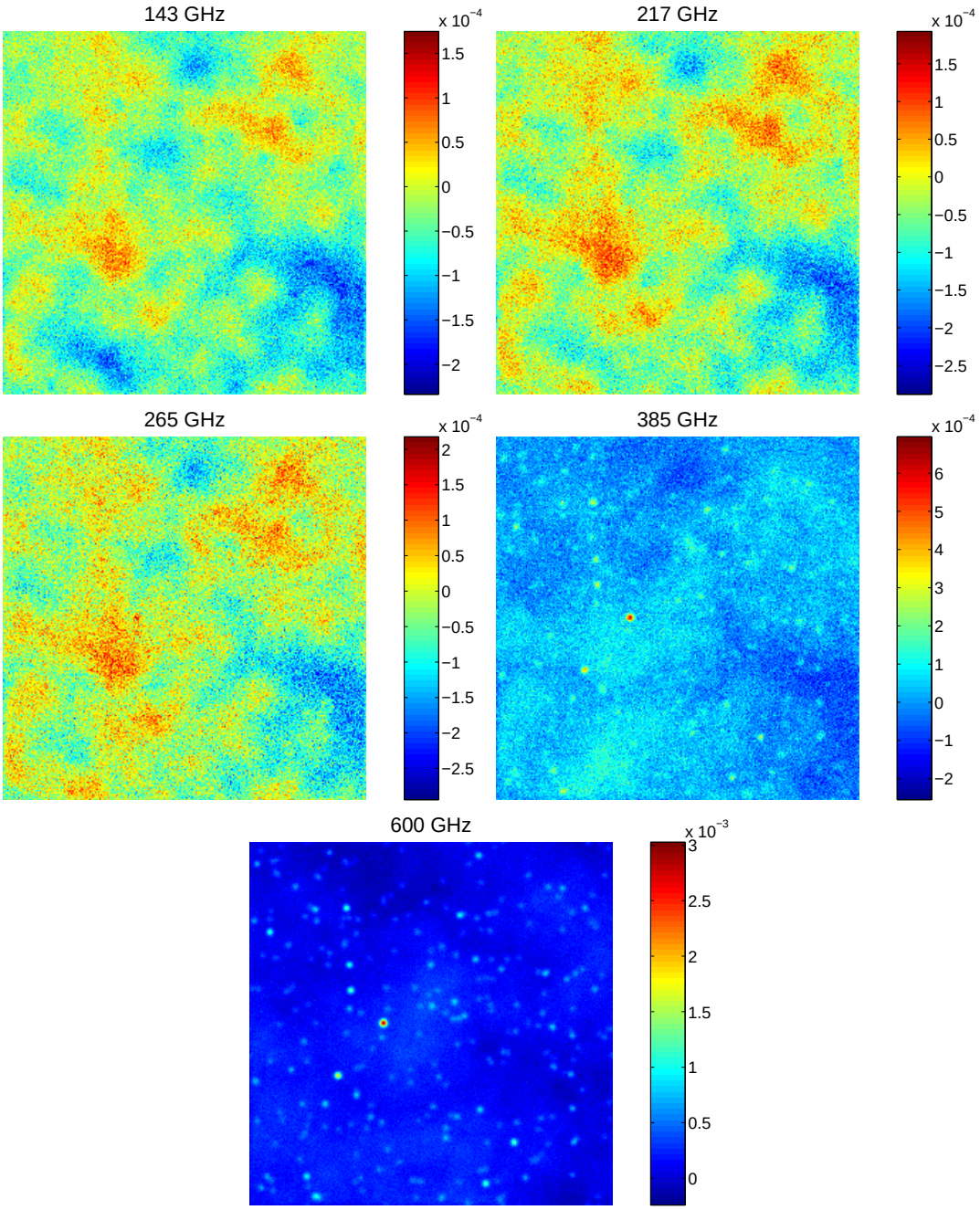

FIG. 5.5 – Simulated 3.2 by 3.2 degrees observed maps at different frequencies of observation. Top left : 143 GHz, top right : 217 GHz, middle left : 265 GHz, middle right : 385 GHz, bottom : 600 GHz.

GHz. In 2006, the Atacama Cosmology Telescope (ACT) will begin an SZ survey of galaxy clusters exactly in this range, with three frequencies of observations : 145 GHz, 217 GHz and 265 GHz. The arrow labeled with the letter "i" points at the location of a point source in Figure 5.4, showing that even at these well selected frequencies, very bright point sources do appear. In [28], the authors quantified potential bias in the reconstruction of the SZ signal due to the point sources under the conditions of this experiment. For other experiment such as the Planck mission, larger frequencies (300-600 GHz) will be observed too, giving the possibility to extract point sources better.

The picture would not be complete without taking into account the beam size at different frequencies and the noise. Figure 5.5 displays maps corresponding to those of Figure 5.4, when noise is added and beams of the correct frequency-dependent size are used. (The beam and noise parameters correspond to those of the experiment described in Section 5.5).

5.2 How to quantify the results ?

A standard measure of the residual error between two images is the Root Mean Square (RMS) error : $\mathrm{RMS}(I_1, I_2) = \sqrt{\frac{1}{N}\sum_{x,y}\left[I_1(x,y) - I_2(x,y)\right]^2}$ where N is the total number of pixels in the images. The RMS error corresponds to the L^2 norm of the difference between the images and is therefore a global measure. The RMS error can be computed at each scale of a wavelet decomposition (or of another decomposition), thus exhibiting at which spatial length the two images are more similar or different. We find that for the Cosmic Microwave Background and the Galaxy dust maps, the RMS error in pixel space and the RMS error computed by scale, combined with visual inspection of the maps and residuals give a sufficient idea of the quality of our reconstructed maps. Indeed, these signals are spread across the whole sky so that a global measure of error treating each pixel the same way gives a good sense of the quality of the reconstructions. The point sources and the clusters' signals, on the other hand, have to be quantified by other means because they are made of intense and compact objects surrounded by void. The RMS error, whether in pixel space or by scale, sums up the contributions from all locations in space, giving a poor idea of how localized the signals are.

Point sources

The principal features of point sources are their brightness, their sparseness and the fact that their extent is smaller than the pixel size. The reconstructed maps of point sources we obtain are rather conservative, and never yield the reconstruction of a point source where it did not exist. However the maps may be polluted by low intensity signal which is either white noise and residual of the galaxy dust map. These low intensity pollutants are rather easy to separate from the estimated point sources by thresholding the reconstructed map. Thus, to asses the quality of a reconstructed point source map, we first examine the level of low intensity residual. The quality of

the estimated point sources is then defined by the number of point sources identified, the extent of each compact object in the reconstructed map that corresponds to a point source and the average fraction of the true intensity of the point sources recovered.

Clusters of galaxies

As for the clusters of galaxies, the task is a little more complicated because clusters vary dramatically in size, shape and magnitude. Moreover, the clusters are the main focus of our study, so we need to define carefully how to asses the quality of these maps. Clusters are compact objects with a peak of intensity at the center, and are distributed sparsely across the sky. Our strategy to detect them in a map is to isolate local maxima that are global maxima over a small fixed angle θ_1. This corresponds roughly to deciding that the size of the smallest cluster we want to detect is θ_1. The order of magnitude of θ_1 is then the typical size of a cluster, i.e. a few arcminutes. The exact value of θ_1 has to be adjusted to the resolution of the data at hand. We refer to the local maximum as the "center" of a detected cluster.

The studies we present here use simulated data, therefore we can compare the reconstructed maps to the ground truth. To do so, we apply the detection procedure described above to both the "true" and the reconstructed map. A reconstructed cluster is then considered as a true detection if its center is closer than a predefined angle θ_2 to the center of a cluster in the original map. In some rare cases, the reconstructed map shows several local maxima (of different intensity) even though there is only one "true" cluster. In this case, we take only the most salient maxima to make our quantified quality assessment. The purity of a sample of reconstructed clusters is then defined as the fraction of clusters in this sample that are true detections.

Our next task is to determine which observable is the most reliable to derive cosmological parameters. Because of the convolution by the beam and the different sizes of the clusters, it is likely that the maximal or central value of the y-parameter is not reliably restored in the reconstructed maps. Instead, we expect that averaged values are more reliable. Again, the angle θ_3 over which the y-parameter should be averaged to find a relevant observable for the clusters has to be tailored to the experiment at hand. We assess how well the collection of reconstructed average y-parameters matches the "true" values by linear regression : we fit a line through the cloud of point formed by the pairs $(y_{true}, y_{reconstructed})$ in two dimensions. The slope of this line tells us what the bias is in the averaged y-parameter of the reconstructed maps compared to the true value. That is to say, if we detect a cluster in the reconstructed map, with averaged y-parameter value $y_{reconstructed}$, we predict that the true corresponding averaged y-parameter value is $y_{predicted} = \frac{y_{reconstructed}}{s}$, where s is the slope. We define the spread Δ of this cloud of points by the average departure from the best fitting line, rescaled to the true value, i.e. :

$$\Delta = E\left\{ \frac{|y_{true} - y_{predicted}|}{y_{true}} \right\} = E\left\{ \frac{|y_{true} - \frac{y_{reconstructed}}{s}|}{y_{true}} \right\} \tag{5.3}$$

The slope and spread then give us a way to take into account the bias in the re-

constructed map when we predict the number of "true" clusters above a predefined average y-parameter. The ratio between the number of such clusters predicted from the reconstructed map to the actual number of such clusters in the original map is called the completeness.

With these tools to asses the quality of our reconstructions, we can now explain the analysis of the performances of the methods we proposed in Chapter 2 and 3 in the context of three different experiments. Each of the next three sections of this chapter is devoted to the description of one experiment and the corresponding results. Before we go on, let us make two remarks. Firstly, the tools we just have presented use the fact that we know the original clusters map. This is a way to benchmark the performances of our algorithms, however these tools would have to be further developed in the case of real data. Secondly, we quantify general aspects of the clusters' reconstruction, such as the number of clusters and their intensity, leaving for later the quantitative study of finer properties, such as their shape and the structures surrounding the peak of intensity in a cluster. We nevertheless examine these finer properties qualitatively.

5.3 ACT : a high resolution experiment

The ACT experiment is a ground-based survey that will collect data on a 100 degree square area of the sky. ACT stands for the Atacama Cosmology Telescope. This telescope is designed specifically for high-sensitivity large-area surveys of the sky requiring dedicated observations for months at a time. It is located in Chile and the experiment is planned to start in November 2006. The ACT survey will map the Cosmic Microwave Background anisotropies from angular scales of a degree down to an arcminute. One of the goals of this survey is to find and study all galaxy clusters in the portion of sky imaged that have a mass greater than 3.10^{14} solar masses through their Sunyaev-Zel'dovich effect. Data will be acquired at 145, 217 and 265 GHz, the expected beam size and noise level are given in Table 5.1.

ACT experiment

Frequency of observation ν (GHz)	Beam size fwhm (arcmin)	Noise level $\sigma(\mu$K)
145	1.7	2
217	1.1	3.3
265	0.93	4.7

Tab. 5.1 – The characteristics of the ACT experiment. The RMS detector noise per full-width-half-maximum pixel, labeled σ, is given in thermodynamic temperature units.

As we pointed out in Subsection 5.1.4, the CMB and SZ signals are largely dominant at these frequencies. The contribution of the Galaxy dust is negligible and this component can be safely disregarded. Point sources may cause some problems, as was pointed out in [28], however, we choose to leave them out because they are

not so troublesome at the frequencies for ACT. As a consequence, we do not assess here the quality of the reconstruction of very compact clusters, i.e. clusters smaller than the beam size which is one arcminute, because they may in practice be confused with the point sources. Since most massive clusters are larger than the beam, it is expected that a great number of theses clusters will be resolved. Moreover, at this resolution, clusters appear aspherical (see Figure 5.10), and a challenge will be to also detect and resolve the outskirts of massive clusters. With these goals in mind, we assess the quality of the reconstruction methods proposed in Chapter 2 and Chapter 3 by using simulations containing the contribution of the CMB and the SZ signals only at the frequencies and with the beam size and noise specified in table 5.1. The CMB is simulated as a Gaussian random field using a power spectrum derived from the best-fitting WMAP parameters [5]. The SZ simulated maps are obtained from hydrodynamical simulations by Zhang et al. [64]. We analyze 24 sets of simulations, each of which covers a 1.44 square degree area of the sky. Our study then covers roughly a one third of the area that will be covered by the true ACT experiment.

To get a rough idea of the level of the noise compared to the contribution of the CMB and SZ signals in the observations, we display in Figure 5.6 the power spectrum of each signal at 145 (left panel) and 265 GHz (right panel). The power spectrum of the CMB and SZ signals are modulated by their frequency dependence. The SZ signal dominates the CMB at scales coarser than 3 arcminutes. The spectra of the CMB and SZ signals have to be multiplied by the beam spectrum to obtain the spectral contribution in the observation. Since the noise level is moderate and the beam size quite small, the SZ signal is dominant over the noise for scales coarser than two arcminutes (resp. one arcminute) at 145 (resp. 265) GHz. Therefore, we do expect that the reconstruction of the SZ will be accurate at least down to the beam size (one arcminute).

We used both our statistical and functional methods to analyze these data. We compare four sets of results : the Gaussian, profile and truncated profile prior distributions for the SZ clusters and our best variational results, using an weighted L^2 norm in wavelet space for the CMB and a Besov norm for the clusters. (The CMB prior is fixed to Gaussian for the statistical method). These different methods were explained, respectively, in Section 3.4 and 2.5.2.

5.3.1 Reconstructions of the Cosmic Microwave Background

Figure 5.8 shows a typical 1.2 by 1.2 degree CMB map (top panel) together with the reconstruction obtained from each algorithm. The corresponding residual maps are in the following figure (Fig. 5.9). Visual inspection of these figures suggests that the four methods considered yield reconstructions of the CMB maps of the same quality. We computed the average over the 24 simulations considered of the RMS in pixel space and scale by scale. The RMS in pixel space is 1.12×10^{-6} for all methods. The RMS per scale are plotted in Figure 5.7.

Both the residuals maps of Figure 5.9 and the RMS per scale in Figure 5.7 show that the most errors occur at the 4.4 arcminutes scale, which corresponds to extended clusters. We notice on Figure 5.7 that the distribution of the error per scale is slightly

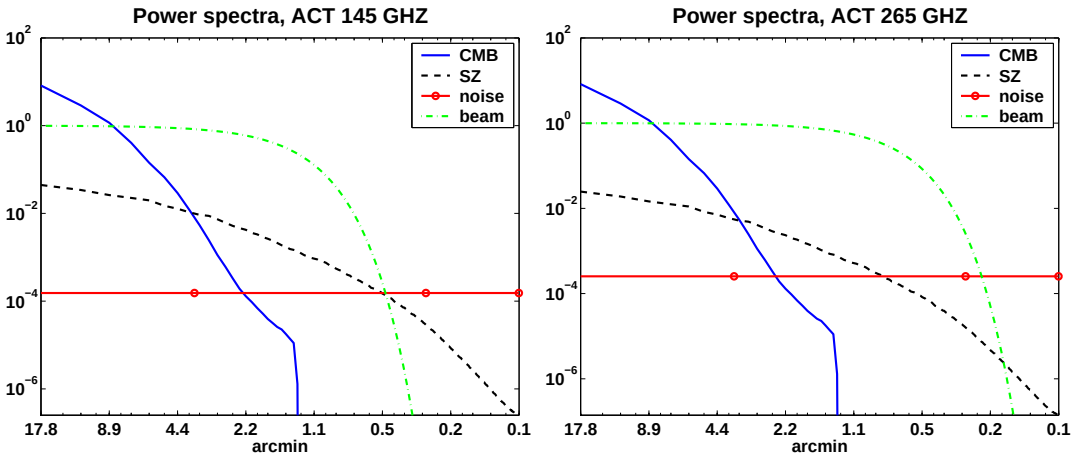

FIG. 5.6 – Power spectra of the signals contributing to the observation for the ACT experiment. Left : at 145 GHz, right : at 265 GHz. The horizontal axis indicates the inverse of the spatial frequency (on a logarithmic scale), so that small numbers correspond to fine spatial scales and large numbers to coarse spatial scales.

different for the functional algorithm than for the statistical ones. The functional method seems to reconstruct more accurately larger scale than 8.9 arcminutes while the statistical method performs better at smaller scales. The better accuracy at fine scales for the statistical method may be explained by the use of the neighborhoods which make the estimates more local for the statistical approach than for the functional approach.

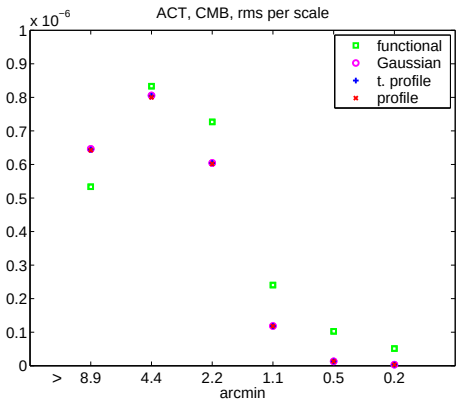

FIG. 5.7 – RMS error in the CMB reconstruction, scale by scale. The results of the Gaussian, profile and truncated profile (noted t. profile) prior lay on top of each other.

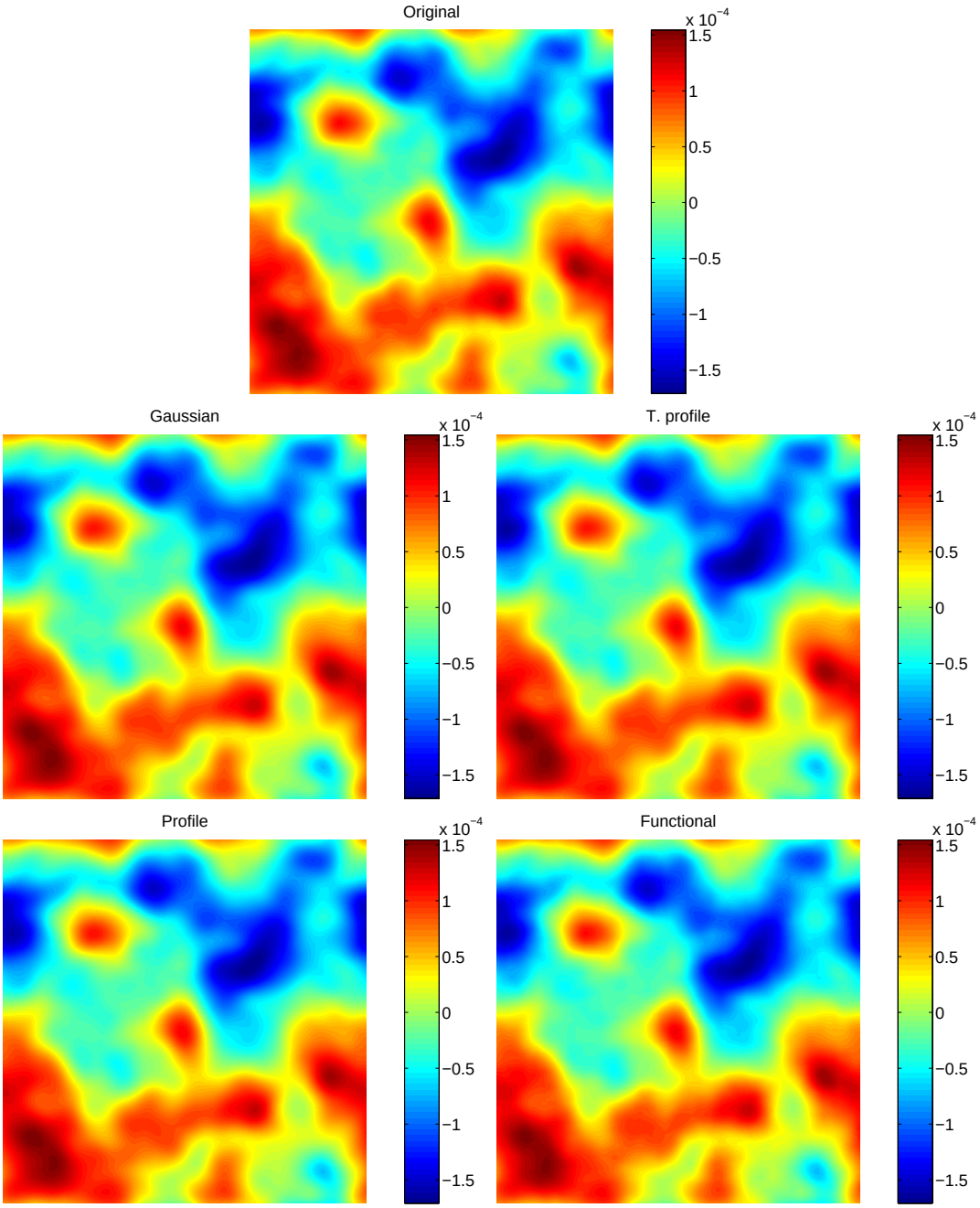

FIG. 5.8 – ACT experiment : CMB. Top : original simulation, other panels : reconstructions. Middle left : Gaussian, middle right : truncated profile, bottom left : profile, bottom right : functional. The maps are 1.2×1.2 degrees.

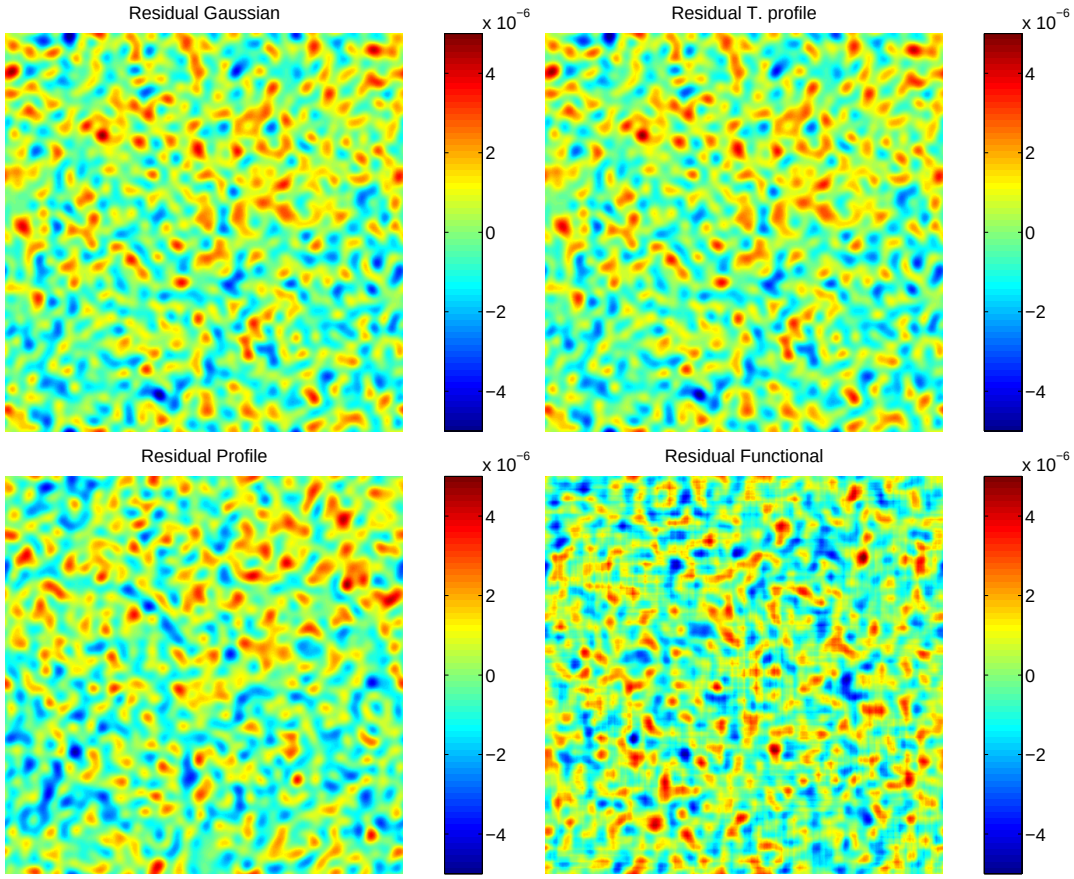

FIG. 5.9 – ACT experiment : CMB residuals. Top left : Gaussian, top right : truncated profile, bottom left : profile, bottom right : functional. The maps are 1.2×1.2 degrees.

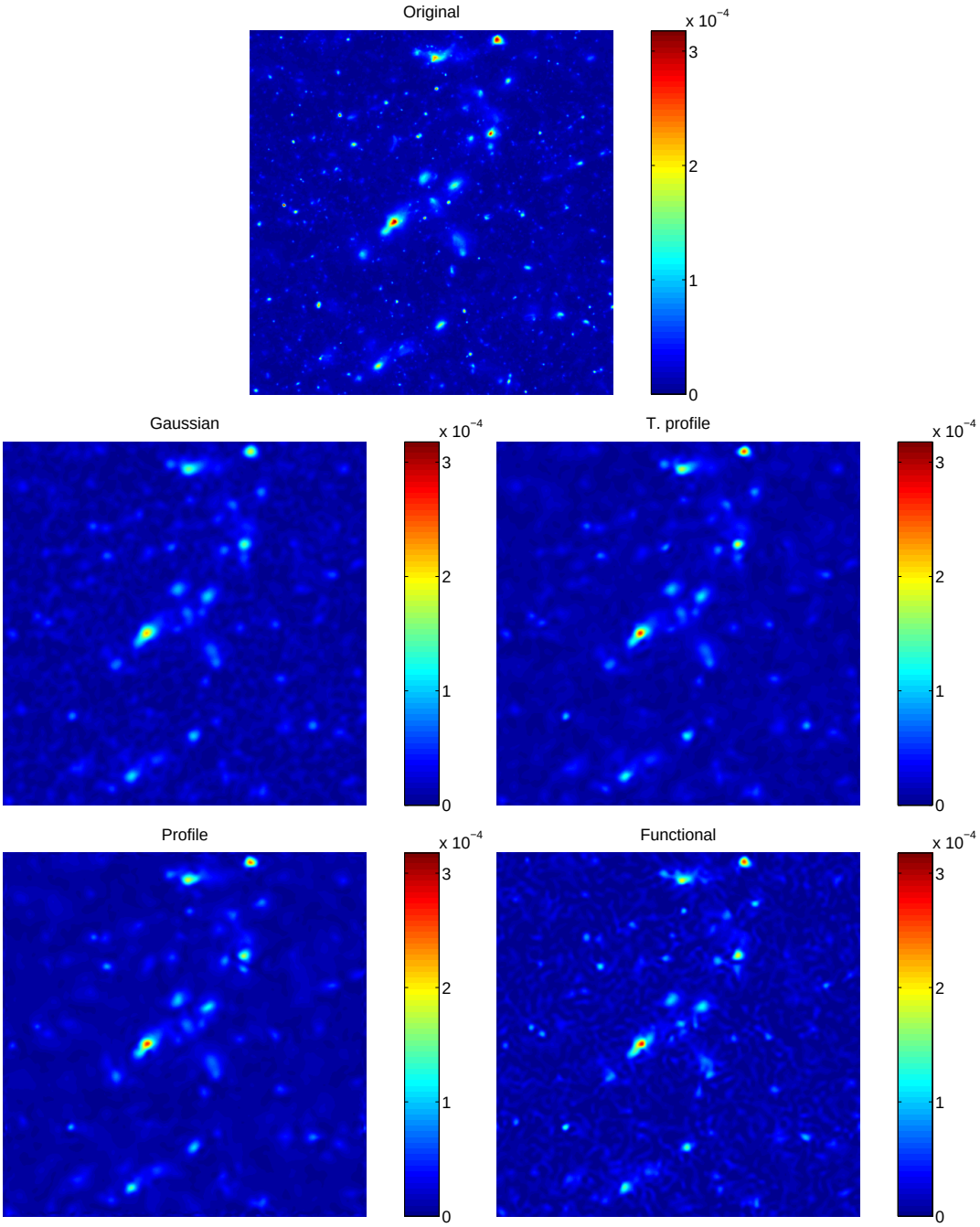

FIG. 5.10 – ACT experiment : SZ clusters. Top original simulation, other panels : reconstructions. Middle left : Gaussian, middle right : truncated profile, bottom left : profile, bottom right : functional. The maps are 1.2×1.2 degrees.

5.3.2 Reconstruction of the SZ clusters

A global quantification of the accuracy of the reconstructed SZ maps is the computation of the average RMS errors in pixel space and per scale for the 24 simulations we used. The RMS error for the different reconstructions are similar. The RMS error in pixel is 8×10^{-6} for the functional method and the Gaussian prior, and 7.7×10^{-6} for the profile and truncated profile priors. The RMS errors per scale are provided in Figure 5.11 and show the same dichotomy, with the functional method and Gaussian prior having a slightly larger RMS error at all scales than the profile priors. Most errors occur at the one arcminute scale, which is the scale of the beam.

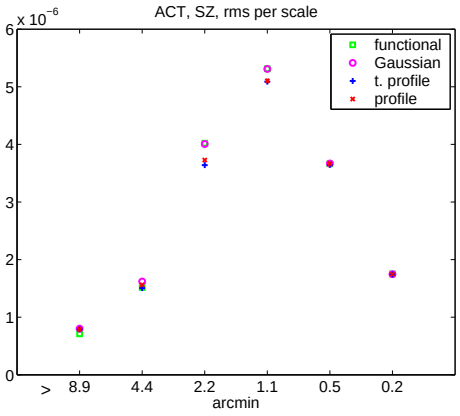

FIG. 5.11 – RMS error in the SZ reconstruction, scale by scale.

The RMS error, per scale or in pixel space, is however not a good indicator of the quality of the reconstructed maps when quality denotes relevance for deriving astrophysical constraints. We illustrate this fact by showing a simulated 1.2 by 1.2 degree map together with the reconstructed maps of our four methods in Figure 5.10. The qualitative comments we can make from visual inspection of such maps are consistent with the quantitative study that follows.

Qualitative inspection of the reconstructed maps.

Visual inspection of the reconstructed maps tells us that the Gaussian prior underestimates the central value of the most intense clusters, whereas the non-Gaussian priors and the functional method perform this task much better. The functional method resolves more compact clusters better than the three statistical methods but on the other hand does a poor job at reconstructing the structures in the outskirts of extended clusters. The Besov norm we chose to constrain the smoothness of the clusters for the functional algorithm promotes local fast transitions and is therefore able to pick up 89 % of the central intensity of bright clusters (we explain in the next subsection how this number is computed). However, the background in the functional reconstruction (see bottom right panel of figure 5.10), shows that structures of lower intensity reconstructed with this method are rather elongated. As a result the

outskirts of the clusters are not well resolved and it is difficult to assess the extent of a cluster using this method. The statistical method, on the other hand, is able to link together smoothly the outskirts of the clusters because it takes into account the correlations between neighboring wavelet coefficients. The use of the profile prior for the statistical method induces a substantial improvement in the reconstruction of the central y-parameter of a cluster compared to the Gaussian prior. However, lower intensity clusters are better resolved under the Gaussian prior because it imposes less regularity in the low-intensity range than the profile prior. Suspecting that our deconvolution method for the prior tends to overweight low values of the multiplier, we truncated the profile prior. The results obtained with this second profile (middle right panel of Figure 5.10) show a compromise between the initial profile and the Gaussian prior : the central parameter of bright clusters is as good as for the profile prior and lower intensity structures are better reconstructed.

Quantitative inspection of the reconstructed maps.

When it comes to infer cosmological parameters from number counts in other wavebands (i.e. X-ray or optical), the common practice is to retain only the brightest clusters which are less affected by selection effects and have a better characterized scaling function. We adopt here the same strategy with SZ clusters, also motivated by the fact that they are less affected by reconstruction errors.

Our first task is to determine which observable is the most reliable to derive cosmological parameters. As we explained in the previous section we have to select the angle θ_c over which the y-parameter should be averaged in the context of this experiment. We smoothed the original and reconstructed maps over angles ranging from 0 to 1.8 arcminutes, which is the size of the largest beam. For each such angle we compute the slope and spread associated to the best fitting line to the clouds of points defined by the original versus reconstructed averaged y-parameter for each detected cluster in the original map. Increasing the value of the averaging angle, we find a big improvement when the angle reaches 0.9 arcminute, which corresponds to the smallest beam of the experiment. The left panel of figure 5.12 shows the evolution of the slope and spread with the averaging angle for the fifty brightest clusters in this study. The slope and spread improve further after the 0.9 arcminute angle ; however, because the most compact clusters are about 1 arcminute wide, smoothing over larger angles will blend the background with the clusters' y-parameter values unevenly for compact versus more extended clusters. Therefore, we define our best observable for the ACT experiment to be the y-parameter value averaged over an angle of 0.9 arcminute. The $(y_{true}\ ,\ y_{reconstructed})$ pairs obtained at this angle for the fifty brightest clusters in this experiment are displayed in the right panel of Figure 5.12 for the four reconstruction methods we consider. The top line is the line of perfect reconstruction, while the other lines are the best fitting lines for each reconstruction. The bottom plain line corresponds to the Gaussian prior, the dotted line to the truncated profile, the dash-dotted line to the profile and finally the dashed line is the best fitting line for the functional method. The slope and spread are summarized in table 5.2. We find that the the functional method yields the best slope and spread, reconstructing on average

89% of the intensity of bright clusters with a spread under 10%. The performances of both non-Gaussian statistical methods are comparable although slightly lower, with a slope around 0.84 and spread of 11%. The Gaussian prior performs less well, consistent with what we observed on the reconstructed maps. It is able to recover 69% of the intensity with a spread of 16%.

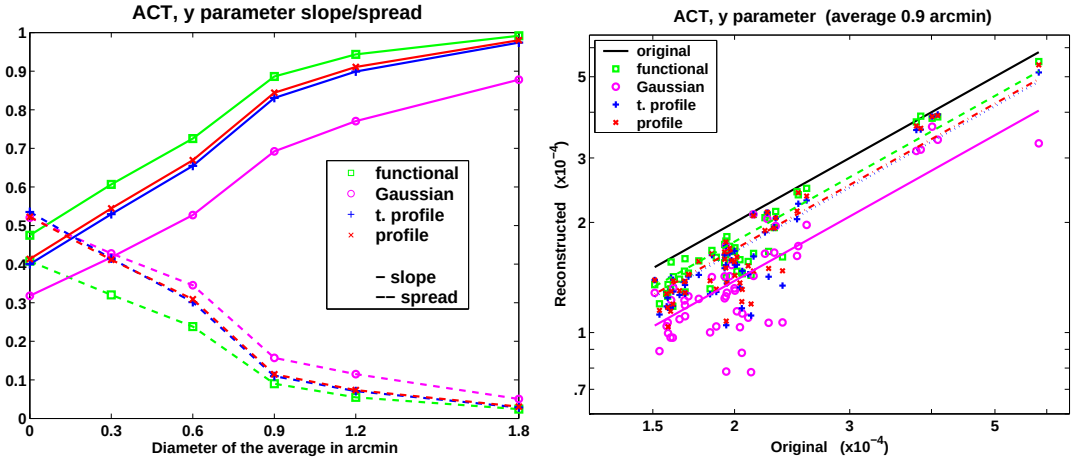

FIG. 5.12 – Left : Slope and spread in function of the averaging angle (labeled diameter). Right : reconstructed versus original central y-parameter averaged at 0.9 arcmin for the fifty brightest clusters.

Method	Statistical			Functional
	Gaussian	Truncated profile	Profile	
Slope	0.69	0.83	0.84	0.89
Spread	0.16	0.11	0.11	0.09

TAB. 5.2 – ACT experiment : slope and spread for the average y-parameter of the 50 brightest clusters.

We finish the quantitative study of the reconstructed maps for the ACT experiment by assessing the quality of predictions that would be made from the reconstructed maps. Two questions come to mind : do the structures found in the reconstructed map really correspond to clusters in the input map ? Can we associate a given threshold in the reconstructed map to an input cluster intensity with high confidence ? To answer these questions, we compute the purity and completeness of the samples for given output intensities. The reconstructed and original maps are smoothed to 0.9 arcmin and clusters are detected in each map. The purity of a sample of reconstructed clusters is the fraction of these clusters that have a counterpart in the original within a radius of 0.6 arcminutes. For a fixed threshold t in the original map, we use the slope s and spread Δ defined earlier to find the sample of detected clusters in each

reconstructed map that would predict true clusters above t. More precisely, we consider that the detected clusters in the reconstructed map above threshold $t.(1-\Delta).s$ predict the number of true clusters above threshold t. The different samples in the reconstructed map then give predictions for the number of true clusters of intensity greater than or equal to a predefined value. Their purity can be compared. We find that reconstructed cluster samples that predict the existence of true clusters of averaged y-parameter above 1.5×10^{-4} are pure, that is, all such detected clusters in the reconstructed map correspond to true clusters. The purity of the statistical maps seems a bit lower than the purity of the functional map as the threshold decreases (see Figure 5.13, left panel). This is consistent with the fact that the corresponding intensity in the reconstructed maps is lower (because the slope is smaller). The completeness is defined as the ratio between the number of clusters in the reconstructed sample to the number of true clusters above the corresponding threshold. The completeness plot in Figure 5.13 shows that the using the threshold $t.(1-\Delta).s$ in the reconstructed maps is too optimistic for the Gaussian and the functional method but yields accurate number counts for the two non-Gaussian statistical priors. We conclude that the non-Gaussian statistical methods predict with great accuracy the number of clusters of averaged y-parameter above 1.5×10^{-4}, with no false positive. In this study, we found 50 such clusters, thus the real ACT experiment will detect around 150 such clusters. This is an appropriate number count to derive cosmological constraints.

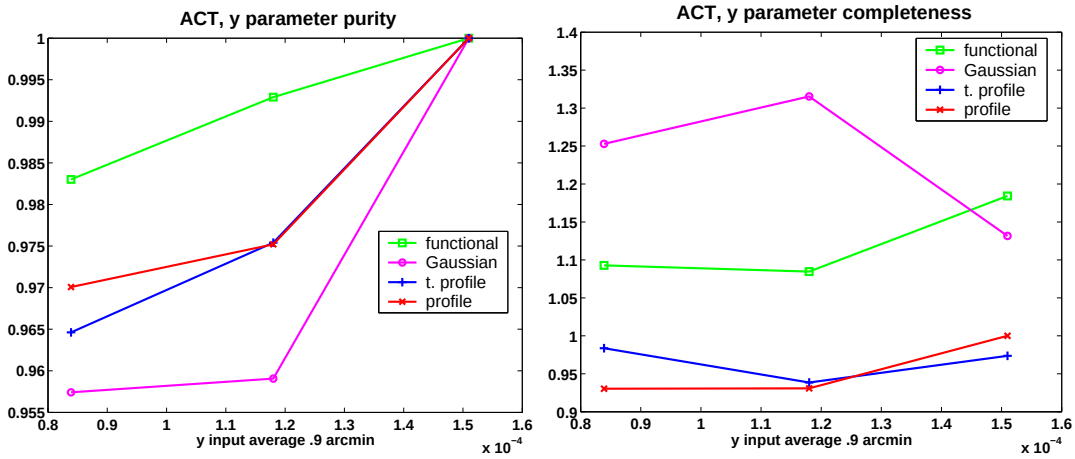

FIG. 5.13 – Purity (left) and completeness (right) of the reconstructed samples.

5.4 Planck : a lower resolution experiment

The Planck mission is designed to image the anisotropies of the Cosmic Microwave Background Radiation over the whole sky. Although it will give unprecedented sensitivity and angular resolution for such a task, the beam sizes and level of noise

are noticeably bigger than for the ACT experiment (see Table 5.3). The size of the smallest beam, around 5 arcminutes, is quite large compared to the typical cluster size (1 to 10 arcminutes).

Planck experiment

Frequency of observation ν (GHz)	Beam size fwhm (arcmin)	Noise level $\sigma(\mu\text{K})$
143	7.1	6
217	5.0	13
353	5.0	40

TAB. 5.3 – The characteristics of the Planck experiment at the frequencies used in this work. The RMS detector noise per full-width-half-maximum pixel, labeled σ, is given in thermodynamic temperature units.

In this work, we assess the quality of our reconstruction methods on simulated observed maps containing only the CMB and SZ clusters' contribution. We consider the three frequencies of observations where the contributions of these two signals are the strongest : 143, 217 and 353 GHz. The actual Planck experiment will make measurements at higher frequencies, where point sources and galaxy dust are dominant. We rely on the fact that the use of these observations will allow to locate and estimate point sources, and focus on the CMB and SZ signals. We use ten simulations, each of which is a 10 by 10 degrees map. The CMB maps are simulated by Gaussian random fields using a power spectrum derived from the best-fitting WMAP parameters [5] (same as for the ACT experiment described in the previous section). The SZ simulated maps are taken from White [62, 65].

In Figure 5.14, we display the power spectrum of the different signals contributing to the observations at the frequencies where the clusters' signal is the strongest. The power spectrum of the CMB and of the clusters is scaled by their frequency dependence, however the convolution by the beam is not taken into account. As expected, the CMB signal dominates the SZ clusters' signal except at fine scales (around 2 arcminutes). Taking into account the convolution by the beam, i.e. multiplying the power spectrum of the CMB and SZ clusters' signal by this of the beam, one can see from these plots that the noise dominates the SZ signal at most scales. Under these conditions, we expect Planck to detect the most massive (or extended) clusters only. The large area covered by the experiment, however, will allow to detect a sizable number of them.

We used both our statistical and functional methods to analyze these data. Similarly to our study of the previous experiment, we compare four sets of results : the Gaussian, profile and truncated profile prior distributions for the SZ clusters and our best variational results, using an weighted L^2 norm in wavelet space for the CMB and a Besov norm for the clusters. We refer to Section 3.4 and 2.5.2 for the details of each method.

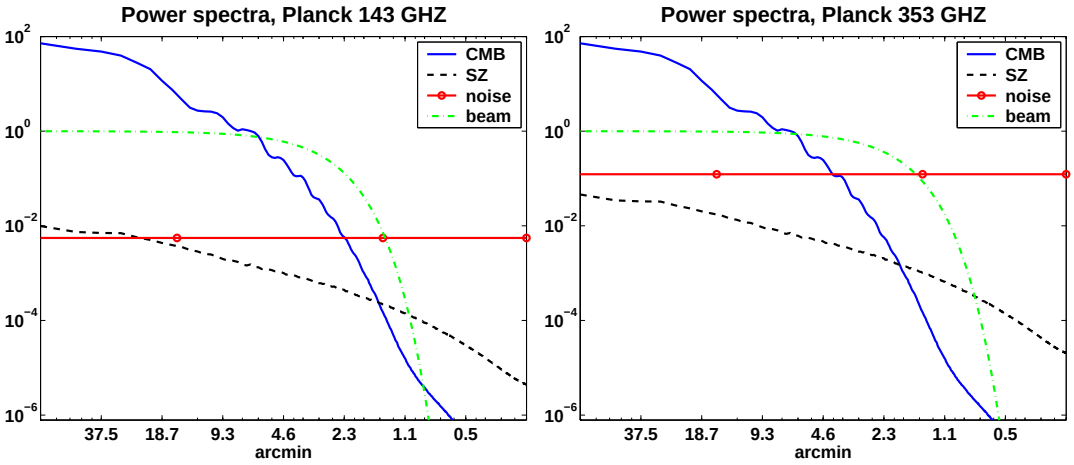

FIG. 5.14 – Power spectra of the signals contributing to the observation for the Planck experiment. Left : at 143 GHz, right : at 353 GHz.

5.4.1 Reconstructions of the Cosmic Microwave Background

As is the case for the ACT experiment, the quality of the reconstructions of the Cosmic Microwave Background is similar for the four methods. Figure 5.16 shows a 5 by 5 degrees portion of one of the simulated maps together with the reconstruction obtained from each method. The total RMS error for the statistical reconstructions is slightly lower (1.12×10^{-5}) than for the functional method (1.16×10^{-5}). This difference of precision is spread across all scales (see the RMS per scale plots in Figure 5.15).

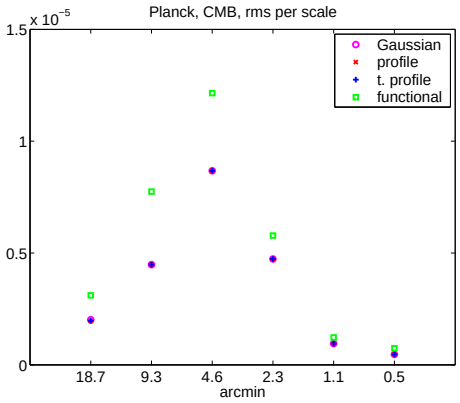

FIG. 5.15 – RMS error in the CMB reconstruction, scale by scale.

Both the RMS per scale plots and the residual maps of Figure 5.17 tell us that the reconstructions are accurate for scales larger that the typical beam size (5 arcminutes). The size of the beam in this experiment is the limiting factor of the reconstructions

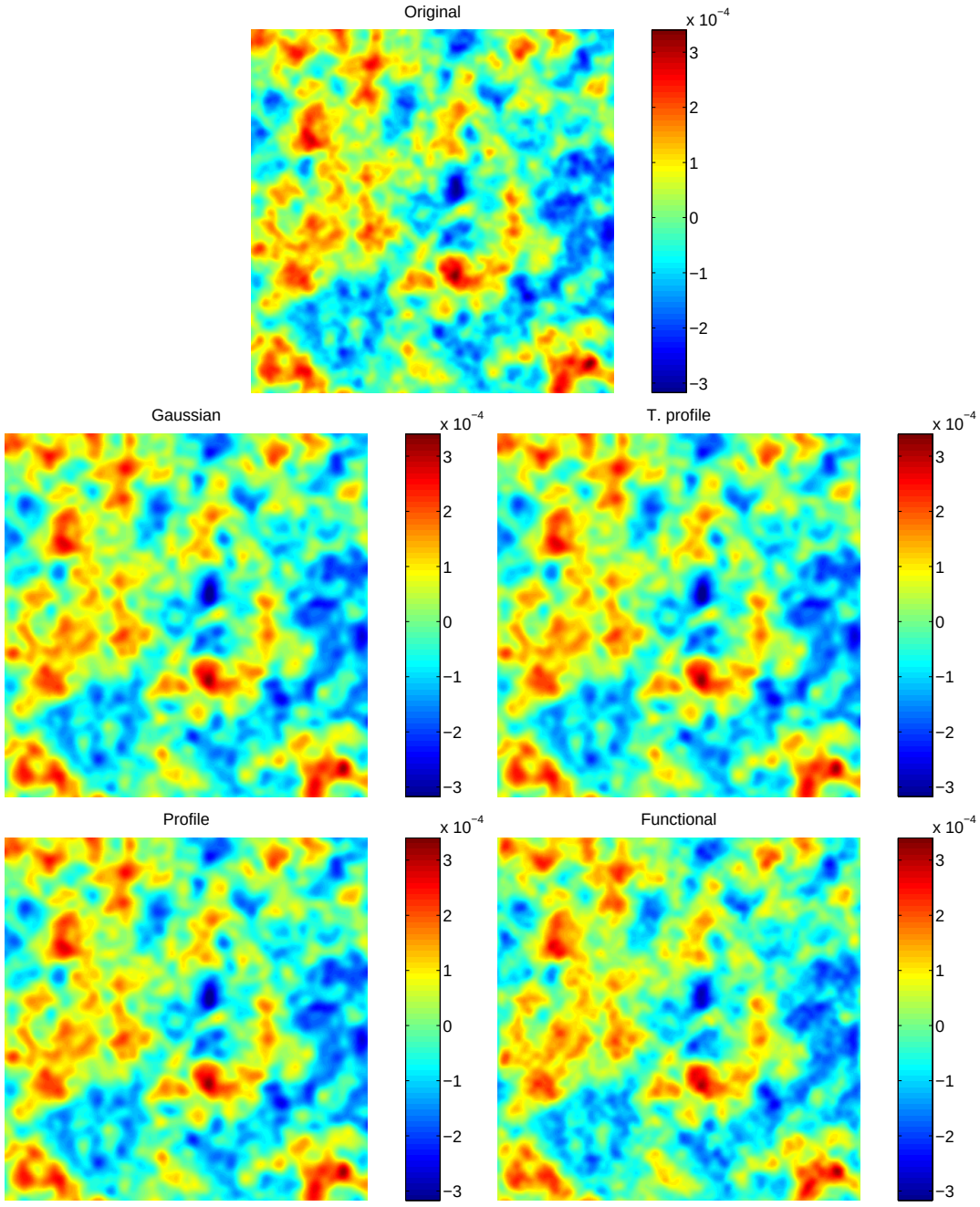

FIG. 5.16 – CMB, Planck experiment. Top : original simulation, other panels : reconstructions. Middle left : Gaussian, middle right : truncated profile, bottom left : profile, bottom right : functional. The maps are 5×5 degrees.

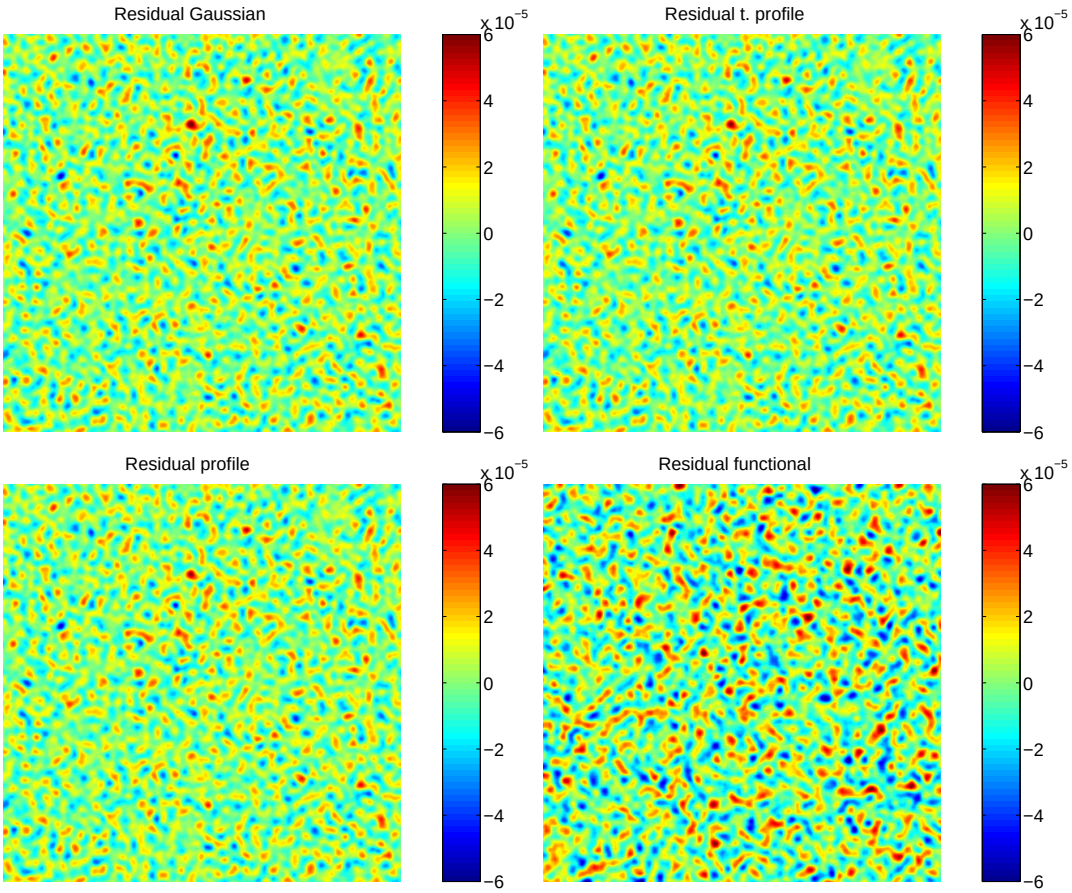

FIG. 5.17 – CMB, Planck experiment : residuals. Top left : Gaussian, top right : truncated profile, bottom left : profile, bottom right : functional. The maps are 5×5 degrees.

of the Cosmic Microwave Background fluctuations, regardless of which method is employed.

5.4.2 Reconstruction of the SZ clusters

As expected from the size of the beam and the level of noise in this experiment, we find that we can reliably reconstruct only bright and extended clusters. In figure 5.18, we show an input y map together with the reconstructed maps for each method. In these figures we see that the statistical and functional methods have very different behavior at low signal-to-noise ratio. The statistical method is rather conservative, yielding a low amplitude reconstruction, even for massive and bright clusters, whereas the functional method allows to recover the amplitude of the signal better at the expense of having a strong residual signal spread across the map. The maps obtained by the statistical method on the contrary are well localized. We notice the effect of the prior distribution is the same as for the ACT experiment. The Gaussian assumption for the clusters allows to recover more low intensity signal. The profile prior causes the amplitude of the bright clusters to be better reconstructed, but at the same time underestimates lower clusters. The truncated profile prior reaches a consensus between the two. Only a few clusters can be detected from the reconstructed statistical maps (low completeness), however, the purity is maximal : every cluster detected (above a threshold y-parameter of 2×10^{-5}) is a true cluster. This is not the case for the functional method. Because of the rather intense residual structure, a significant number of clusters would be detected in the functional map that do not exist. One would need to increase the threshold up to 5×10^{-5} to obtain maximal purity in this case.

We selected the eight brightest and most extended clusters out of our ten simulations to quantitatively compare the reconstruction of the central y-parameter with the different methods. Typically, these massive clusters are about 10 arcmin wide and their maximal y-parameter exceeds 5×10^{-5}. As is the case for ACT experiment, we find that the observable that reaches the best trade-off between the adequation to the original data and the spread is the average value of the central y-parameter over an angle of roughly the same size as the beam. Figure 5.19 shows the output averaged central y-parameter found in the reconstructed maps versus input averaged central y-parameter in the original maps for the eight clusters selected. The top line is the line of perfect reconstruction, the other lines show the best fitting line for each method. In the table 5.4, the slope and spread corresponding to these eight clusters is quoted for each reconstruction.

As can be observed on the reconstructed maps in Figure 5.18, taking in account the non-Gaussianity improves the reconstruction of the central y-parameter by a factor 4 (truncated profile) to 6 (profile) over the Gaussian prior in the statistical method. The functional method is even more accurate, improving the reconstructed values by a factor 9 over the Gaussian statistical method and 1.5 compared to the best statistical method. Although the slope is significantly improved over the Gaussian prior, the spread in the non-Gaussian statistical reconstructions is somewhat high : around 30% of the nominal value. This could be a potential problem when it comes to deriving

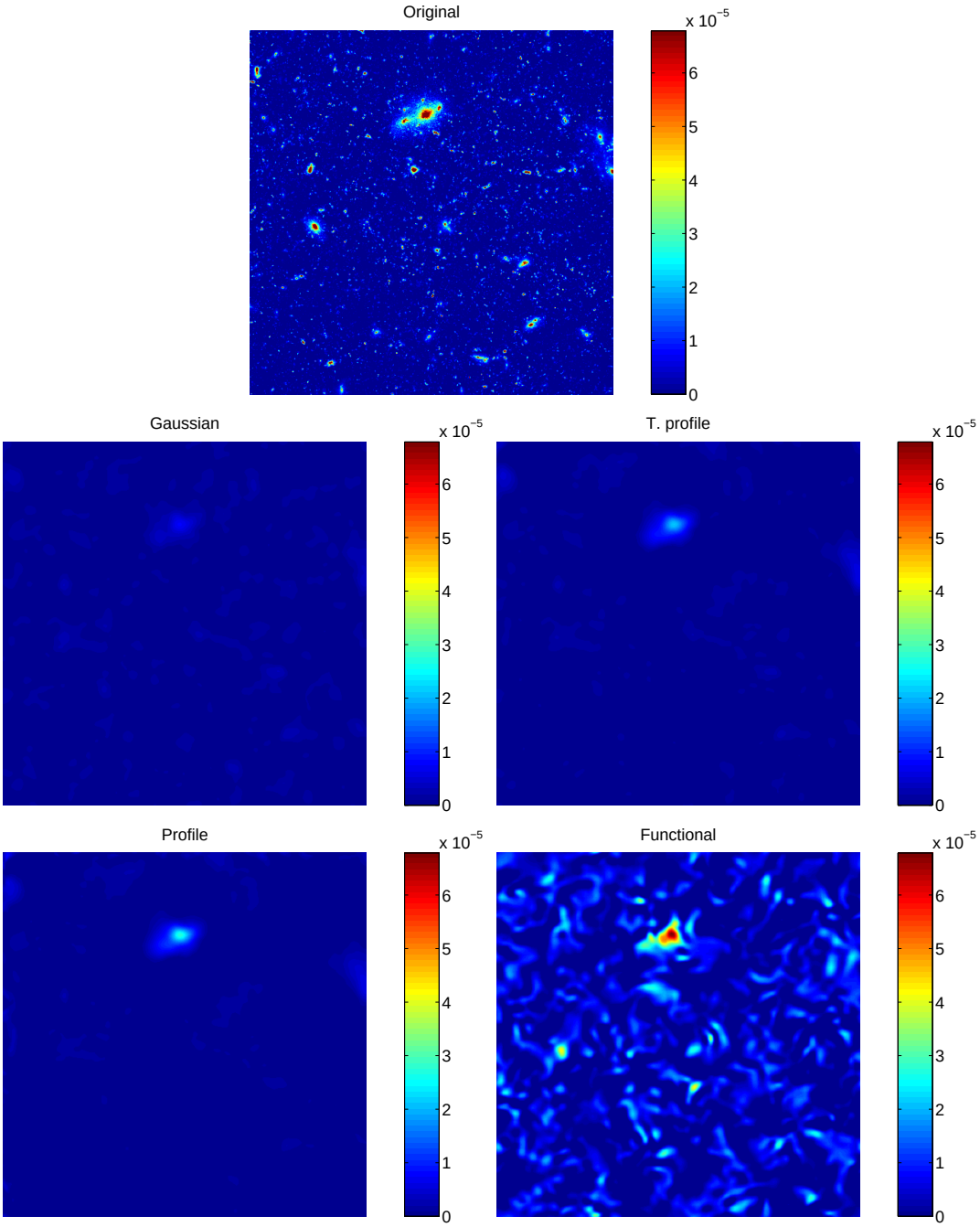

FIG. 5.18 – SZ clusters, Planck experiment. Top original simulation, other panels : reconstructions. Middle left : Gaussian, middle right : truncated profile, bottom left : profile, bottom right : functional. The maps are 5×5 degrees.

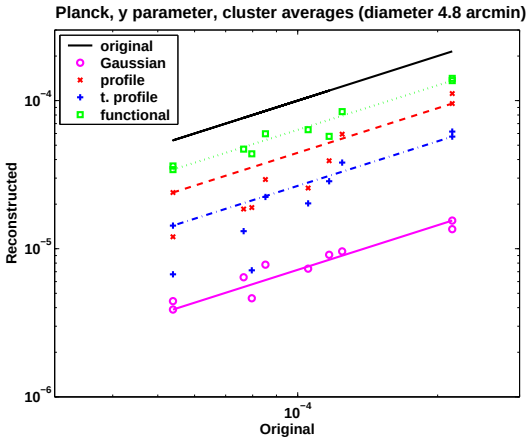

FIG. 5.19 – Reconstructed versus original central y-parameter (4.8 arcmin average).

Method	Statistical			Functional
	Gaussian	Truncated profile	Profile	
Slope	0.07	0.26	0.44	0.63
Spread	0.13	0.27	0.32	0.09

TAB. 5.4 – Planck experiment : slope and spread for the average y-parameter of the eight most massive and bright clusters.

cosmological parameters from these reconstructions. In this regard, the functional method yields a significant improvement over the statistical method altogether, it recovers on average 63 % of the input y-parameter value with a spread that is less than 10 % of this input value.

We conclude that under the conditions of the Planck experiment presented here, only bright extended clusters may be recovered. The two methods we propose complement each other : the shape and localization of the clusters is much better resolved by the statistical method, whereas the functional method is more accurate and reliable for the estimation of the central y-parameter. Neither method seems to be self-sufficient in this case to derive cosmological parameters accurately. However, if one is willing to do the reconstructions with both methods, one could use a map reconstructed from the statistical method to locate massive clusters, (which can not be done reliably with the functional reconstruction,) and then use the result of the functional method to infer the y-parameter of the detected clusters.

The actual performances of the Planck instrument may be better than the ones used in these simulations. In particular the noise in the sky will not be uniformly distributed because some areas will be better sampled than others. We assessed the relevance of the noise level on the performances of the statistical method by performing a similar analysis on the Planck maps with a reduced level of noise (a factor 7 lower). We find that in these conditions the non-Gaussian statistical methods recover around 60% of the y-parameter with a spread of the order of 10% (see [45] for more details). This shows that the limiting factor for the statistical method in this experiment is the noise level.

5.5 The influence of point sources

In the studies we presented in the last two sections, we have made the simplifying assumption that the contribution of the point sources and the Galaxy dust were negligible or had been extracted from the observed maps before we process them. The third study we present here aims at assessing whether the methods we propose are robust to the presence of the point sources and Galaxy dust. The data we use were simulated by astrophysicist Dominique Yvon and collaborators at CEA, France. The frequencies of observation, beam size and noise level correspond to those of the OLIMPO survey and are given in table 5.5. OLIMPO is an ongoing project which aims at measuring the Sunyaev-Zeldovich effect in many clusters of galaxies during a long-duration balloon flight. The size of the beam in this experiment is intermediate between these of the ACT and Planck experiments we described earlier. The experiment will collect data in four different frequency channels. Examples of observed maps can be found in Figure 5.5 (the 265 GHz observation has been produced for displaying purposes only and is not used in the study). At the two largest frequencies, 385 and 600 GHz, the point sources and Galaxy dust dominate the observations. The CMB signal on the other hand dominates the observation at the lower frequencies, 143 and 217 GHz. The clusters' contribution is maximal at 385 GHz but is largely dominated by point sources and dust, therefore the most reliable channel to observe

the SZ effect is the lowest frequency channel : 143 GHz. The simulated data we study here cover a four hundred degree square portion of the sky.

OLIMPO experiment

Frequency of observation ν (GHz)	Beam size fwhm (arcmin)	Noise level $\sigma(\mu\mathrm{K})/\sqrt{Hz}$
143	3	150
217	2	200
385	2	500
600	2	5000

Tab. 5.5 – The characteristics of the OLIMPO experiment.

5.5.1 Results obtained with the statistical method

For the statistical method, we compare the reconstructions yielded by different sets of distributions. The histograms of the wavelet coefficients of the Galaxy dust are well-fitted by a Gaussian. Moreover, we do not expect that the presence of dust will cause a major deterioration of the clusters' signal, because the Galaxy dust is smooth and slowly varying and fills up the space. Therefore, the prior for the CMB and the Galaxy dust are fixed to Gaussian, and we focus on the influence of different priors for the point sources and clusters.

To get an idea of the problems encountered with the introduction of point sources, we first tried the simplest prior for the clusters, i.e. the Gaussian prior, and compared the results obtained when the point sources are assumed Gaussian to the results obtained using Jeffrey's prior. The Gaussian prior is obviously not the best fitting prior for the point sources because their extent is under a pixel size and they are sparsely distributed. We saw that modeling the non-Gaussianity of the clusters leads to better reconstruction of the SZ effect in the context of the ACT and Planck experiments as well. However, we also found in the two previous studies that the quality of the reconstructions of the CMB signal does not change between the case where the clusters' prior is Gaussian and when it is not. This shows that in the simplified case where only the CMB and SZ effect are present in the observations, the reconstruction of one particular component (the CMB) is largely independent of the prior chosen for the other component (the SZ signal). So the rationale for examining the case where all four priors are assumed Gaussian, even if we know this model is too simple, is to understand whether the reconstructions of the different signals are independent from each other as was the case for the CMB/SZ experiments.

We find that the statistical method is very robust to the introduction of point sources and Galaxy dust as far as the estimation of the CMB and clusters signals are concerned. Indeed, even when all signals are assumed Gaussian, the precision of the reconstructed maps of the CMB and clusters signal is similar to the quality that would be expected from our study of the ACT and Planck experiments. The

CMB signal is very well estimated down to scales around 5 arcminutes, which is slightly larger than the beam size and no traces of point sources or Galaxy dust can be found. The algorithm is able to separate point sources from SZ clusters, and the reconstructed clusters maps have similar quality to those seen for ACT, given the size of the beam. Here the observable we use to assess the quality of the clusters map is the average y-parameter over an angle of two arcminutes. Clusters are detected as local maxima that dominate over a three arcminutes angle and are considered to correspond to a cluster in the original map if the two centers are less than two and a half arcminutes apart. Even when clusters are assumed Gaussian, the purity of the clusters sample from the reconstructed maps is high (about 97 %) for intense clusters (i.e. with central average y-parameter bigger than 10^{-5}). This proves that no intense point sources are confused with the clusters, even when the point sources are modeled with the Gaussian prior.

Surprisingly, the reconstructed map of the point sources allows to locate them accurately, even when they are assumed Gaussian. The estimated point sources are not as compact as a pixel but are extended to roughly the size of the beam. The beam is small enough compared to the mean distance between two point sources that this is not a problem in this experiment. However, the intensity of the point sources is underestimated (around 25 % of their value). Moreover we find the algorithm confused background noise with the point sources map. A white noise is spread out in the reconstructed point source map, but fortunately, its level is lower than the intensity of most point sources. The estimation of the Galaxy dust map is accurate a coarse scale (around 20 arcminutes) but smaller fluctuations are not reconstructed at all.

We now compare the results we obtained by fixing the prior to Gaussian for all signals to the reconstructions obtained when Jeffrey's prior (i.e. the log-uniform distribution on the multiplier) is used for the point sources (still using the Gaussian prior for all the other signals). As expected the point sources map is much better reconstructed, the background noise observed earlier has disappeared. The point sources themselves are still extended to the size of the beam. Their intensity is slightly better estimated than before but is still low (around 35%). Although the prior on the Galaxy dust map has not changed, smaller scales are reconstructed with this set of priors, indicating that the quality of the reconstruction of the Galaxy dust depends on the accuracy of the point sources map. This seems natural since point sources and Galaxy dust have very similar frequency dependence at the frequencies of observation used here. On the other hand, the quality of the CMB and intense clusters' reconstructions remains the same, indicating that the statistical method used here is able to separate signals primarily on the basis of their frequency dependence.

Finally, we studied in further detail the quality of the SZ clusters reconstructions in this experiment by allowing the prior of this signal to be non-Gaussian. The results we obtain are consistent we our remarks above : the reconstruction of other signals is not affected by changing the prior of the clusters. The qualitative and quantitative differences between the Gaussian, the profile, and truncated profile prior are similar to those we found in the ACT experiment. That is to say, the profile prior allows to recover the intense clusters more accurately than the Gaussian prior, at the expense

of underestimating lower intensity clusters and the truncated profile prior reaches a compromise between the other two.

We conclude that under the conditions of the OLIMPO experiment, the presence of the point sources and Galaxy dust will not affect the quality of the SZ maps estimated by using the statistical method we propose.

5.5.2 Results obtained with the functional method

The functional variational method we propose to reconstruct the signals is much more affected by the introduction of point sources. We did not find a balance between the eight terms in the functional (four error terms and four regularization terms) that allows to accurately recover all signals at the same time. With the nominal values described in Section 2.5.2, the CMB is reconstructed correctly although it is a little smoother than expected, but only a coarse scale approximation of the clusters' signal is recovered. The point sources maps is very well localized (the extent of estimated point sources is typically smaller than the beam size). However, only 35% of their intensity is recovered in the estimated point sources map, and the remainder of this signal is attributed to the Galaxy dust, in the form of extended point sources of the size of the beam on top of the Galaxy dust itself.

This lead us to conduct a smaller case study in order to determine whether the Galaxy dust and point sources can be separated at all using this method. We generated observations with the parameters of the OLIMPO survey, only omitting the contribution of the CMB and SZ cluster's signal. From these observations we tried to separate the Galaxy dust signal from the point sources. We find that the regularizing terms of these two signals have to be balanced taking into account the relative amplitude of the Galaxy dust variations and the intensity of the point sources. This leads to choosing the parameters γ_4 and γ_3 so that $\gamma_4 \sum_{\lambda=(j,k)\in\Lambda} 2^{3j} |\langle f_4, \varphi_\lambda \rangle|^2 \sim 100\ \gamma_3 \sum_{pixel} |f_3(pixel)|$, rather than of the same order. With these parameters, the functional algorithm is able to reconstruct both the point sources and the Galaxy dust with great accuracy. In particular, the estimated point source map is free of noise and the intensity of the point sources is recovered at 90%. Moreover, the extent of the estimated point sources is extremely close to one pixel, with the intensity decaying sharply at the four closest pixels if it is not zero. Such accuracy in the point sources map can not be achieved by the statistical method because it is constrained to estimate the point sources map in wavelet space, causing the extent of the point sources to be limited by the finer wavelet scale.

However, we find that the balance between point sources and Galaxy dust terms is greatly affected by the reintroduction of the CMB and clusters signal. In particular, a complicated interplay occurs between the reconstructions of the clusters signal, the Galaxy dust signal and the point sources. As a result, the estimation of the clusters' map is either too coarse or contains point sources that will make the detection of clusters unreliable. Finding a better way to balance the different terms is extremely difficult because contrarily to what we observed for the statistical method, the estimation of one particular signal is greatly affected by the estimation of the other signals, making it impossible to study the influence of one parameter at a time.

We conclude from this study that the presence of point sources is a major concern with the functional variational algorithm we proposed, preventing the method to reconstruct accurately all signals at the same time. However, we find that in the restricted case where only the point sources and the Galaxy dust maps are to be extracted, this method is able to locate and estimate the point sources with great accuracy both in intensity and in spatial extent. Therefore, the functional algorithm we propose could be used in other type of experiments where the focus is the point sources, to locate and estimate them accurately. From a more general point of view, the success of the restricted experiment containing only point sources and Galaxy dust shows that our innovative use of norms defined by different tight frames for different signals is promising.

5.6 Summary of the results

In this chapter, we have applied both the variational approach and the statistical approach we described in Chapters 2 and 3 to estimate the major astrophysical components present in surveys of the sky at the frequencies between 100 and 600 GHz. There are four of these components : the Cosmic Microwave Background, the Sunyaev-Zeldovich effect, the infrared point sources and the Galaxy dust. Our goal is to obtain reliable information on the clusters of galaxies by reconstructing accurate maps of the Sunyaev-Zeldovich effect.

Since the SZ effect is a fluctuation of the CMB radiation, the reconstruction of the CMB radiation is inherent to the estimation of the clusters of galaxies through their Sunyaev-Zeldovich signature. The point sources and Galaxy dust, however can be seen as pollutants of a second order. They dominate larger frequencies of observations while the CMB and clusters signal are more intense at smaller frequencies. Therefore, we first assessed the quality of our methods on simulated data ignoring point sources and Galaxy dust. Since different sky survey may have very different resolution, noise level and be able to cover different extent of the sky, we studied two test cases of different nature. The first one, ACT will cover a small portion of the sky with a resolution of the order of one arcminute and moderate noise level. The second experiment we consider, Planck, will cover the whole sky with a resolution of five arcminutes and higher level of noise. In a third study, with intermediate resolution and moderate noise, we assessed the influence of point sources and Galaxy dust.

For each experiment, we compared the results obtained for the functional method to several sets of results obtained with the statistical method, where different priors were used. The “Gaussian statistical approach” refers to the case where the clusters’ signal is modeled by a Gaussian prior and the “non-Gaussian statistical approach” to other cases.

Our findings are the following :

– The most reliable observable of the SZ clusters is the y-parameter averaged over an angle of the same order as the beam size. (The y-parameter is the quantity intrinsic to a cluster of galaxies that determines the amplitude of the resulting Sunyaev-Zeldovich effect).

– In the absence of points sources and Galaxy dust, both methods perform similarly. The CMB signal is reconstructed accurately down to the scale of the smallest beam. However some differences are noticed :
  The functional approach and non-Gaussian statistical approach outperform the Gaussian approach in the estimation of intense clusters. Moreover, the statistical method does a better job at estimating the structure of the clusters whereas the functional approach recovers more intensity.
  For the high resolution experiment, ACT, we find that the clusters' signal is very accurately estimated by both methods, especially for the intense clusters. We conclude that both the non-Gaussian statistical reconstructions and the functional reconstruction yield estimates of the average y-parameter of intense clusters that could be used to constrain cosmological quantities.
  For the low resolution experiment, Planck, we find that the reconstructions of the SZ effect are limited to bright and very extended clusters. The reliability of the detection of these clusters in the functional reconstructions is low because large residual structures appear. However, the estimation of the averaged y-parameter is remarkably stable at the location of the true clusters. This, in a sense, completes the performances of the non-Gaussian statistical approach. In that case, extended clusters can be detected reliably because the structure surrounding the peak of intensity are well estimated. However the spread of the average y-parameter reconstructed is too high to be trusted. We conclude that under these conditions neither methods are self-sufficient to derive cosmological parameters from the reconstructed SZ maps. However, we determined that the limiting factor in this case is the noise level, which may be improved in the true experiment in some areas of the sky that are observed for a longer time.
– The statistical method is robust to the introduction of point sources and Galaxy dust, leading to accurate estimates of the CMB and clusters signal. We determined that for this approach, the estimation of a single component does not affect other components which have a different frequency dependence. Thus, it is not necessary with this method to recover the point sources accurately to obtain a satisfying clusters' signal.
  This is not the case for the functional approach, where a complicated interplay between the different terms makes it difficult to study the precision of the reconstruction of each component separately. As a result, we were not able to recover all four signals simultaneously with this approach in order to find a satisfying cluster map. We note however that the functional approach we propose can be used to recover the point sources with almost perfect accuracy both in terms of their intensity and their spatial extent, when the number of signals is reduced.

Bibliographie

[1] D. F. Andrews, C. L. Mallows, *Scale mixtures of normal distributions*, Journal of the Royal Statistical Society, Series B (Methodological), Vol. 36, No.1 (1974), 99-102.

[2] S. Anthoine, E. Pierpaoli, I. Daubechies, *Deux méthodes de déconvolution de mélanges de composantes ; application à la reconstruction des amas de galaxies*, accepted to GRETSI'05.

[3] R. A. Battye, J. Weller, *Constraining cosmological parameters using Sunyaev-Zel'dovich cluster surveys*, 2003, Phys. Rev. D., **68**, 083506.

[4] H. H. Bauschke, P. L. Combettes, and S. Reich, *The asymptotic behavior of the composition of two resolvents*, Nonlinear Analysis : Theory, Methods, and Applications, vol. 60, no. 2, pp. 283-301, January 2005.

[5] Bennett, C. L. ; Halpern, M. ; Hinshaw, G. ; Jarosik, N. ; Kogut, A. ; Limon, M. ; Meyer, S. S. ; Page, L. ; Spergel, D. N. ; Tucker, G. S. ; Wollack, E. ; Wright, E. L. ; Barnes, C. ; Greason, M. R. ; Hill, R. S. ; Komatsu, E. ; Nolta, M. R. ; Odegard, N. ; Peiris, H. V. ; Verde, L. ; Weiland, J. L., *First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations : Preliminary Maps and Basic Results*, The Astrophysical Journal Supplement Series, Volume 148, Issue 1, pp. 1-27.

[6] A.J. Bell and T.J. Sejnowski, *Fast blind separation based on information theory*, Proc. Intern. Symp. on Nonlinear Theory and Applications, Las Vegas, Dec. 1995.

[7] J. Cardoso, H. Snoussi, J. Delabrouille, and G. Patanchon, *Blind separation of noisy Gaussian stationary sources*. Application to cosmic microwave background imaging, in Proc. EUSIPCO Vol. 1, pp 561-564, 2002.

[8] A. Chambolle, R. A. DeVore, N.-Y. Lee and B. J. Lucier, *Nonlinear Wavelet Image Processing : Variational Problems, Compression, and Noise Removal through Wavelet Shrinkage.* IEEE Trans. Image Processing **7** (1998), 319–335.

[9] A. Chambolle and P.L. Lions, *Image recovery via total variation minimisation and related problems*, Numerische Mathematik, vol. 76, no. 2, pp.167–188, 1997.

[10] A.Cohen, Y.Meyer and F.Oru, *Improved Sobolev inequalities*, Proc. Séminaires X-EDP, No.IV, Centre Math., Ecole Polytechnique, Palaiseau, France, 1998.

[11] A. Cohen, I. Daubechies, and J.C. Feauveau, *Biorthogonal bases of compactly supported wavelets*, Comm. Pure & Appl. Math., 45, pp. 485–560, 1992.

[12] R. Coifman and D. Donoho, *Translation invariant de-noising*, in Lecture Notes in Statistics : Wavelets and Statistics, vol. New York : Springer-Verlag, pp. 125–150, 1995.

[13] R. R. Coifman, Y. Meyer, and M. V. Wickerhauser, *Size properties of wavelet-packets*. In Wavelets and their applications, pp. 453–470. Jones and Bartlett, Boston, MA, 1992.

[14] P. L. Combettes and V. R. Wajs, *Theoretical analysis of some regularized image denoising methods*, Proceedings of the IEEE International Conference on Image Processing, vol. 1, pp. 321-324. Singapore, October 24-27, 2004.

[15] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, *Wavelet-based statistical signal processing using hidden Markov models*, IEEE Trans. Signal Processing, vol. 46, pp. 886-902, Apr. 1998.

[16] I. Daubechies, M. Defrise, C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 2004, vol. 57, issue 11, p. 1413–1457.

[17] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Lecture Notes nr. 61, SIAM , 1992.

[18] I. Daubechies and G. Teschke, *Wavelet-Based Image Decompositions by Variational Functionals*, Proc. SPIE Vol. 5266, p. 94-105, Wavelet Applications in Industrial Processing ; Frederic Truchetet ; Ed., Feb. 2004.

[19] I. Daubechies and G. Teschke, *Variational image restoration by means of wavelets : simultaneous decomposition, deblurring and denoising*, accepted for publication in Applied and Computational Harmonic Analysis, 2005.

[20] C. De Mol and M. Defrise, *Inverse imaging with mixed penalties*, Proceedings URSI EMTS 2004, Ed. PLUS Univ. Pisa, pp. 798-800, 2004.

[21] D. L. Donoho, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition*, App. and Comp. Harmonic Analysis, vol. 2, pp. 101-126, 1995.

[22] D. Donoho, *For Most Large Underdetermined Systems of Linear Equations, the minimal l^1-norm near-solution approximates the sparsest near-solution*, technical report, Department of Statistics, Stanford University, 2004.

[23] D. Donoho, *Neighborly Polytopes and Sparse Solutions of Underdetermined Linear Equations*, technical report, Department of Statistics, Stanford University, 2004.

[24] D. L. Donoho and M. Elad, *Maximal Sparsity Representation via l^1 Minimization*, the Proc. Nat. Aca. Sci., Vol. 100, pp. 2197-2202, March 2003.

[25] M. Figueiredo, R. Nowak, *Bayesian wavelet-based signal estimation using non-informative priors*. In Proc. Thirty-Second Asilomar Conf. Signals, Systems, and Comp., Pacific Grove, CA. IEEE Computer Society Press, 1998.

[26] L. Landweber, *An iterative formula for Fredholm integral equations of the first kind.* Am. J. Math. **73** (1951), 615–624.

[27] W. Hu, *Self-consistency and calibration of cluster number count surveys for dark energy* Phys. Rev. D **67**, 081304(R) (2003).

[28] K M. Huffenberger, U. Seljak, *Prospects for ACT : Simulations, power spectrum, and non-Gaussian analysis*, New Astronomy, Volume 10, Issue 6 , June 2005, Pages 491-515

[29] J. Kalifa, S. Mallat, *Thresholding Estimators for Linear Inverse Problems and Deconvolution*, The Annals of Statistics, vol. 31, no. 1, pp 58-109, 2003.

[30] J. Kalifa, S. Mallat, *Deconvolution by Thresholding in Mirror Wavelet Bases*, IEEE trans.on Image Processing, 2003.

[31] N. G. Kingsbury, *Image Processing with Complex Wavelets*, Phil. Trans. Royal Society London A, September 1999, on a Discussion Meeting on "Wavelets : the key to intermittent information ?", London, February 24-25, 1999.

[32] N. G. Kingsbury, *Complex wavelets for shift invariant analysis and filtering of signals*, Journal of Applied and Computational Harmonic Analysis, vol 10, no 3, May 2001, pp. 234-253.

[33] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Analysis Machine Intell. PAMI-6 (1984) 721-741.

[34] A. Jalobeanu, *Models, Bayesian estimation and algorithms for remote sensing data deconvolution*, PhD thesis, Université de Sophia-Antipolis, 2001.

[35] M. Lang, H. Guo, J.E. Odegard, C.S. Burrus, and R.O. Wells, *Nonlinear processing of a shift invariant DWT for noise reduction*, Proc. of SPIE, vol. 2491, 1995, pp. 640-651.

[36] E. S. Levine, A. E. Schulz, M. White, *Future Galaxy Cluster Surveys : The Effect of Theory Uncertainty on Constraining Cosmological Parameters*. The Astrophysical Journal, Volume 577, Issue 2, pp. 569-578.

[37] L. B. Lucy, *An iterative technique for the rectification of observed distributions*, Astronomical Journal, Vol. 79, p. 745 (1974).

[38] S. Majumdar and J. Mohr, *Importance of Cluster Structural Evolution in Using X-Ray and Sunyaev-Zeldovich Effect Galaxy Cluster Surveys to Study Dark Energy*. The Astrophysical Journal, volume 585, part 1 (2003), pages 603-610.

[39] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1998.

[40] S. Mallat, *A theory for multiresolution signal decomposition : The wavelet representation,* IEEE Trans. Pattern Anal. Machine Intell., vol. 11, pp. 674–693, 1989.

[41] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, *Low-complexity image denoising based on statistical modeling of wavelet coefficients*, IEEE Signal Processing Lett., vol. 6, pp. 300-303, Dec. 1999.

[42] J.J. More, *The Levenberg Marquardt algorithm : implementation and theory*, pages 105-116 in Lecture Notes in Mathematics, (Numerical Analysis : proceedings of the biennal conference, Dundee), G. Watson ed., Springer-Verlag,New York, 1978.

[43] F. Murtagh, J.-L. Starck and A. Bijaoui, *Multiresolution in astronomical image processing : a general framework*, International Journal of Imaging Systems and Technology, 6, 332-338, 1995.

[44] F. Oru, *Le rôle des oscillations dans quelques problèmes d'analyse non-linéaire*, PhD thesis, CMLA, ENS-Cachan, France, 1998.

[45] E. Pierpaoli, S. Anthoine, K. Huffenberger, I. Daubechies, *Reconstructing Sunyaev-Zeldovich clusters in future CMB experiments*, Mon. Not. Roy. Astron. Soc. Volume 359, Issue 1, pp. 261-271 - May 2005.

[46] E. Pierpaoli, S. Anthoine, *Finding SZ clusters in the ACBAR maps*, COSPAR 2004, in press.

[47] J. Portilla, V. Strela, M. Wainwright, E. Simoncelli, *Image denoising using a scale mixture of Gaussians in the wavelet domain*, IEEE Trans. Image Proc., 2003, vol. 12, issue 11, p. 1338–1351.

[48] J. Portilla, E. P. Simoncelli, *Image Restoration using Gaussian Scale Mixtures in the Wavelet Domain*, 9th IEEE Int'l Conf on Image Processing. vol. II, pp. 965-968, Barcelona, Spain. September 2003.

[49] W. H. Richardson, *Bayesian-based iterative method of image restoration*, J. Opt. Soc. Am., vol. 62, p. 55-59 (1972).

[50] J. Romberg, H. Choi, and R. Baraniuk, *Bayesian Tree-Structured Image Modeling using Wavelet Domain Hidden Markov Models*, IEEE Transactions on Image Processing, Vol. 10, No. 7, pp. 1056-68, July 2001.

[51] L. Rudin and S.Osher, *Total variation based image restoration with free local constraints*, Proc. IEEE ICIP, Austin-texas, USA, Nov. 1994, vol. I, pp. 31-35.

[52] I. W. Selesnick, *Hilbert transform pairs of wavelet bases*, IEEE Signal Processing Letters, 8(6) :170-173, June 2001.

[53] I. W. Selesnick.*The design of approximate Hilbert transform pairs of wavelet bases*, IEEE Trans. on Signal Processing, 50(5) :1144-1152, May 2002.

[54] L. Sendur and I. W. Selesnick, *Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency*, IEEE Trans. on Signal Processing. 50(11) :2744-2756, November 2002.

[55] J. Shapiro, *Embedded image coding using zerotrees of wavelet coefficients*, IEEE Trans. Signal Processing, vol. 41, pp. 3445-3462, Dec. 1993.

[56] E. P. Simoncelli, W. T. Freeman, E. H. Adelson and D. J. Heeger, *Shiftable Multi-Scale Transforms*, IEEE Trans. Information Theory, Special Issue on Wavelets. Vol. 38, No. 2, pp. 587-607, March 1992

[57] J.-L. Starck, F. Murtagh and A. Bijaoui, *Multiresolution support applied to image filtering and restoration*, Graphical Models and Image Processing, 57, 420-431, 1995.

[58] J.-L Starck and F. Murtagh, *Image restoration with noise suppression using the wavelet transform*, Astronomy and Astrophysics, 288, 342-348, 1994.

[59] Private communication of J-L. Starck to I. Daubechies, Nov 2004.

[60] M. J. Wainwright and E. P. Simoncelli, *Scale mixtures of Gaussians and the statistics of natural images*, Adv. Neural Information Processing Systems, S. A. Solla, T. K. Leen, and K. R. MÃ$\frac{1}{4}$ller, Ed. Cambridge, MA : MIT Press, 2000, vol. 12, pp. 855-861.

[61] M. Tegmark, G. Efstathiou, *A method for subtracting foregrounds from multi-frequency CMB sky maps* MNRAS, 281 :1297-1314, 1996.

[62] M. White, *Studying Clusters with Planck*. The Astrophysical Journal, Volume 597, Issue 2, pp. 650-658.

[63] N. Wiener, *The interpolation, extrapolation, and smoothing of stationary time series*, Wiley, New York, 1949 .

[64] P. Zhang, U.-L. Pen, and B. Wang, *The Sunyaev-Zeldovich Effect : Simulations and Observations*, ApJ, 577, 555, 2002.

[65] http ://pac1.berkeley.edu/tSZ/PlanckSZ/

[66] http ://taco.poly.edu/WaveletSoftware/

[67] http ://www.cns.nyu.edu/ eero/steerpyr/

Titre : Approches en ondelettes pour la séparation et la déconvolution simultanées. Application à des données astrophysiques.

Résumé : Cette thèse est consacrée au problème de séparation de composantes lorsque celles-ci sont des images de structure différente et que l'on en observe un ou plusieurs mélange(s) flou(s) et bruité(s). Les problèmes de déconvolution et de séparation, traditionnellement étudiés séparément, sont ici traités simultanément.

Une façon naturelle d'aborder le problème multicomposants/multiobservations est de généraliser les techniques de déconvolution d'une image unique. Le premier résultat est une étude mathématique d'un tel algorithme. Preuve est faite que celui-ci est convergent mais pas régularisant et une modification restaurant cette propriété est proposée. Le sujet principal est le développement et la comparaison de deux méthodes pour traiter la déconvolution et séparation simultanées de composantes. La première est basée sur les propriétés statistiques locales des composantes tandis que dans la seconde, ces signaux sont décrits par des espaces fonctionnels. Les deux méthodes utilisent des transformées en ondelettes redondantes pour simplifier les données.

Les performances des deux algorithmes sont évaluées et comparées dans le cadre d'un problème astrophysique : l'extraction des amas de galaxies par l'effet Sunyaev-Zel'dovich dans les images multispectrales des anisotropies du fond cosmique. Des simulations réalistes sont étudiées. On montre qu'à haute résolution et niveau de bruit modéré, les deux méthodes permettent d'extraire des cartes d'amas de galaxies de qualité suffisante pour des études cosmologiques. Le niveau de bruit est un facteur limitant à basse résolution et la méthode statistique est robuste à la présence de points sources.

Mots-clés : estimation/détection de signaux, ondelettes, approche statistique/variationnelle

---

Title : Different Wavelet-based Approaches for the Separation of Noisy and Blurred Mixtures of Components. Application to Astrophysical Data.

Abstract : This thesis addresses the problem of separating image components that have different structure, when several observations of blurred mixtures of these components are available. In the image processing literature, the deblurring problem has been well described for a single component in a single image and the separation problem mainly studied without blurring. In this thesis, the full problem is addressed globally, the separation being done simultaneously with the denoising and deblurring of the data, by generalizing methods that exist for the enhancement of a single image.

The first result is a mathematical analysis of a heuristic iterative algorithm for the enhancement of a single image. This algorithm is proved to be convergent but not regularizing ; a modification is introduced that restores this property. The main object of this thesis is to develop and compare two methods for the multi-components/multi-observations problem : the first method uses functional spaces to describe the signals ; the second method models the local statistical properties of the signals. Both methods use wavelet frames to simplify the description of the data.

Both algorithms are evaluated with regards to a particular astrophysical problem : the reconstruction of clusters of galaxies by the extraction of their Sunyaev-Zel'dovich effect in multifrequency measurements of the Cosmic Microwave Background anisotropies. Realistic simulations are studied. It is shown that both methods yield clusters maps of sufficient quality for subsequent cosmological studies when the resolution of the observations is high and the level of noise moderate. Then some limiting factor are pointed out.

Keywords : signal estimation/detection, wavelets, statistical/variational approach.